CAUSAL INFERENCE FROM OBSERVATIONAL STUDIES IN THE PRESENCE OF PARTIAL INTERFERENCE

Brian G. Barkley

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biostatistics in the Gillings School of Global Public Health.

Chapel Hill
2018

Approved by:

Michael G. Hudgens

M. Alan Brookhart

Michael Emch

Michael Kosorok

Donglin Zeng

ABSTRACT

Brian G. Barkley: Causal Inference From Observational Studies in the Presence of Partial Interference
(Under the direction of Michael G. Hudgens)

Analyzing data to estimate the effect of treatment on health outcomes can play a major role in the fields of personal and public health. Interference occurs when the treatment of one individual affects the outcome of another individual. This work aims to develop statistical methodology for inference about causal effects from observational studies in the presence of interference. We assume partial interference throughout: interference may exist within clusters of individuals, but not between distinct clusters. In each paper we propose estimators that are consistent and asymptotically Normal; estimators for the asymptotic variance are also proposed. Finite-sample performance of each estimator is investigated, and each method is illustrated by analyzing a cholera vaccine study in Matlab, Bangladesh.

In the first paper, we propose a method for inverse probability-weighted estimation of target estimands in the presence of partial interference that is more robust to model mis-specification than existing methods. This technique relies on an algorithm which combines machine learning and mixed effects methods to determine the relationship between treatment and covariates assuming a certain correlation structure. We employ the algorithm on a training sample as a data-adaptive model selection procedure. We recover the set of rules that the algorithm uses for prediction to transform the covariates in a testing sample. We proceed by fitting a model to the transformed covariates to estimate propensity scores for IPW estimation of target estimands.

We propose a matching technique for estimating causal effects in the presence of partial interference in the second paper. These estimators extend methods that are commonly employed when no interference is assumed. The proposed methods can be carried out without modeling treatment, and may outperform existing IPW estimators in certain scenarios. Extensions of these estimators are discussed.

In the third paper we propose new causal estimands for observational studies in the presence of partial interference. The proposed estimands describe counterfactual scenarios in which there may be within-cluster dependence in the individual treatment selections. These estimands may be more relevant for public health officials. These estimands are identifiable from observational data with parametric assumptions. Inverse probability-weighted estimators for these estimands are proposed.

ACKNOWLEDGMENTS

This work is a testament to Michael Hudgens' talents as a researcher and a mentor. Michael, thank you for knowing when to be patient with me, and when to challenge me. Your daily commitment to your students is remarkable. You have played a great role in how I have grown as a researcher and a person.

The support of my committee has been invaluable. I thank Michael Kosorok and Donglin Zeng for making time to discuss algorithms even when meetings were already running late. I thank Michael Emch for providing excellent knowledge of infectious disease epidemiology, and for resolute communications even when traveling to faraway lands. Finally, I thank Alan Brookhart for his continued support and mentoring, for encouraging me with new opportunities, and for brewing a very good cup of coffee.

I thank the UNC Biostatistics Department for creating an atmosphere of great scholarship. In particular, I thank Dr. Koch for welcoming me into the BCL community, for significant personal and professional mentoring, and for generous financial support.

I am grateful for the people I have had the opportunity to work with. I thank my friends in the CIWI research group for creating a fascinating place to learn. My good friends Bradley Saul and David Pritchard have provided me with incredible guidance, and have been instrumental to my career.

I thank Ania, Mark, Rachel, and Sam for being my family in Carolina.

I am indebted to my parents for all they have provided me; as well the Barkleys, Bueches, Messerschmidts, and Dev. I thank my brother for being my measuring stick.

Finally and foremost, I thank my wife. Jessica, my professional successes are intertwined with my love and appreciation for you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: LITERATURE REVIEW

1.1 Introduction

An important issue in the study of personal and population health is in determining causal effects of health outcomes, instead of simply finding associations and coincidences. Researchers carefully plan, orchestrate, and analyze the data from studies in attempts to determine what events, therapies, behaviors or interventions make noticeable effects on individuals' health outcomes. Most traditional statistical techniques are not designed to draw causal conclusions from data that do not arise from well-crafted and expensive randomized controlled trials, and doing so can result in greatly erroneous decision-making (Holland 1986). In a widely-known example, analyses of several nonrandomized studies with traditional statistical methods suggested a protective effect of hormone therapy against cardiovascular among women; analysis from randomized clinical trials however proved that the therapy did not decrease the risk of heart disease and contrarily could be considered harmful (Hulley et al. 1998).

It is necessary to develop and employ statistical methods that can address causality from nonrandomized trials (Rubin 1974), which is the focus of this manuscript. Doing so would allow us to address causality from a greater range of nonrandomized trials, saving time, money and lives. In particular this manuscript focuses on observational studies in the presence of interference (Cox 1958), which is when one individual's treatment may affect another individual's outcome. This chapter continues with motivating examples in Section 1.2, followed by a literature review in Sections 1.3 and 1.5, and concludes with a summary and research proposal in Section 1.6. In Chapter 1.6, methods

to reduce the risk of bias and increase robustness to statistical modeling assumptions by employing techniques from machine learning and data mining literature are proposed. A new matching estimator is proposed for the cases in which inverse probability weighting may be undesirable due to unstable weights in Chapter 2.8. Finally, new causal estimands and corresponding estimators are presented in Chapter 3.10 for the analysis of observational studies that account for correlation as well as interference.

1.2 Motivating Examples

1.2.1 Recent introduction of rotavirus vaccines in US

Prior to 2006, nearly one-third of severe acute gastrointestinal episodes among children in the United States were attributed to rotavirus infection (Tate et al. 2009), and is the leading cause of gastroenteritis in young US children (Panozzo et al. 2014). Rotavirus vaccines were made available in 2006 with the potential to reduce risk of rotavirus infection, and subsequently coverage across the United States increased quickly (Tate et al. 2009). One vaccine had been estimated to be up to 90% effective in randomized trials (Tate et al. 2009). An early analysis by Tate et al. (2009) comparing rotavirus activity in the years prior to and after the rotavirus vaccine was made available to the US public provided strong evidence that the vaccines reduced the burden of rotavirus in US children.

However, Centers for Disease Control and Prevention estimates showed that nearly one-third of eligible children near the age of 2 years old had not yet completed a full vaccination as of 2011 (Centers for Disease Control and Prevention (CDC) 2012). Similarly, Panozzo et al. (2013) used a large, nationwide database of electronic health records to estimate that up to one-fifth of US infants had not received at least one dose of a rotavirus vaccine as of 2010. Panozzo et al. (2014) carried out a rigorous analysis using the same database source to estimate that vaccination was up to 90% effective in

preventing rotavirus-related hospitalization for the recipient. The analysis by Panozzo et al. (2014) also showed that there was some amount of added protections conferred onto all children, even those who were not vaccinated, due to the increased amount of vaccination in the entire population or "herd immunity".

The above research highlights the plausibility of interference in the study of the causal effects of rotavirus vaccination on rotavirus infections in US children. The nationwide database of electronic health records used by Panozzo et al. (2013; 2014) contains rich information on relevant outcomes, vaccination statuses, and many potential confounders, and so could be used for a causal analysis using methods appropriate for nonexperimental studies. Since the database also contains geographical information, relating physical distance of children, partial interference may be an appropriate assumption. We intend to introduce new estimators that can be used for estimation of causal effects in the presence of interference that control for a high number of covariates in a robust manner.

1.2.2 Cholera vaccine study in Matlab, Bangladesh

From 1985-1988, over 120,000 women and children in Matlab, Bangladesh were offered potential vaccines in a triple-blind randomized trial, and then followed to study the effectiveness of the drugs in reducing the incidence of cholera (Clemens et al. 1988). These results have been presented by Clemens et al. (1988), Ali et al. (2005; 2009), Perez-Heydrich et al. (2014) among others.

Perez-Heydrich et al. (2014) define an individual as being treated when they have taken 2 doses of one of the versions of vaccine. Summary measures of the risk of cholera incidence among study participants, stratified by level of vaccine coverage within geographical region, or bari are given in Table (1.1) below. As coverage increases within the bari from below 50% to more than 75%, the risk of cholera infection amongst the

unvaccinated individuals falls from above 50% to 31%.

Within-bari vaccine coverage	Number of baris	Infections per 1000 unvaccinated
[0%, 20%]	1659	7.6
(20%,40%]	1528	6.3
(40%, 60%]	2057	4.5
(60%, 100%)	1067	3.2

Table 1.1: Infection rate amongst uninfected decreases as vaccination coverage increases

Although the study did offer the vaccines in a strict randomization setting, women and children chose whether to present themselves for inclusion in the randomized study component. Infection outcomes for the women and children who were eligible for the trial but who did not choose to participate in the trial (and therefore went unvaccinated) were still available. Statistical interference is plausible, and so any analysis of this data should include observations for all eligible individuals (and not simply the trial participants) and also control for this nonrandomized component that likely includes a level of self-selection for treatment. Perez-Heydrich et al. (2014) were the first to formally consider the effects related to interference in the potential outcomes framework in an application of this data. We intend to introduce estimators in the presence of interference that exhibit more stable performance than the existing inverse-probability of treatment weighting methods. We also propose new estimands that may be more relevant for scientific inquiry than existing estimands from Tchetgen Tchetgen and VanderWeele (2012).

1.3 Causal Inference from Observational Studies

1.3.1 The Potential Outcomes Framework

This document considers causal inference from the potential outcomes framework. This framework was pioneered by Neyman (1935) and was refined by Rubin (1974).

Holland (1986) further elucidated the similarities and differences between the potential outcomes framework and traditional associative statistical methods. Generally known as the Rubin Causal Model (Holland 1986, Little and Yau 1998, Frangakis and Rubin 2002), the Neyman-Rubin Causal Model (Pearl 1996, Rubin 2005), or even the Neyman-Rubin-Holland Theory (Brady 2002, Sekhon 2008; 2009) causal effects are defined by differences in potential outcomes that follow treatment.

Consider a study of $i = 1, \ldots, M$ individuals and a binary variable representing treatment status A, where $A_i = 1$ indicates that individual i experienced the treatment, and $A_i = 0$ indicates that she did not. Let Y denote subsequent health status, where $Y_i = 1$ indicates that she experienced an unfavorable outcome, and $Y_i = 0$ indicates favorable outcome. Since Y and A may take on different values for different individuals, they are random variables. In the potential outcomes framework, $Y_i(a)$ is the woman's potential outcome that would have been observed if the woman had experienced treatment a.

In one possible scenario, the woman undergoes treatment $A_i = 1$ and then experiences an antecedent health status $Y_i(1)$. In another possible scenario, the woman instead does not undergo treatment $A_i = 0$ and then experiences an antecedent health status $Y_i(0)$ that is quite possibly different from $Y_i(1)$. The causal effect of treatment versus no treatment for the woman is the difference in her potential outcomes in the two scenarios $Y_i(1) - Y_i(0)$.

However, it is not possible to observe the woman both take treatment and also fail to take treatment, and so it is not possible to know the causal effect $Y_i(1) - Y_i(0)$ of treatment on her health outcome. These multiple possible scenarios, at most one of which is factual and the remainder run counter-to-fact, are often referred to as counterfactuals (Lewis 1974; 2001, Glymour 1986, Morgan and Winship 2014). The key issue in causal inference, often called its fundamental problem (Holland 1986) is that

only one of these counterfactuals may be observed, and so it is not possible to truly know the causal effect of a treatment on any individual's health outcome. Instead, causal effects are estimated from observed data.

The above example satisfies the stable unit treatment value assumption (Rubin 1980) or the individualistic treatment assumption (Manski 2013), or SUTVA. SUTVA implies that any individual's outcome is a function of her own treatment, and no individual's treatment can affect another individual's outcome. A major component SUTVA is the assumption of no interference between units (Cox 1958). VanderWeele (2009) discusses the assumption of no interference

$$Y_i(a,k) = Y_i(a,k') \text{ for all } i \text{ and any } a = 0,1, \ \& \ k \neq k', \tag{1.1}$$

where Cole and Frangakis (2009) define $Y_i(a,k)$ as the potential outcome for an individual i when the individual experiences treatment a "by means of" some condition k. Another component of SUTVA is the assumption of causal consistency (Cole and Frangakis 2009, VanderWeele 2009, Pearl 2010),

$$Y_i(A_i) = Y_i(a) \text{ for all } i \text{ when } a = A_i, \tag{1.2}$$

Cole and Frangakis (2009) define this in words: individual i's observed outcome "is the potential outcome, as a function of intervention, when the intervention is set to the observed exposure." Two related assumptions that are often made, at least implicitly, are that the observed variables are measured without error (Rubin 1974, Edwards et al. 2015) and that the time between antecedent treatment and subsequent outcome is sufficient for the treatment to have an effect, if any, on the outcome (Rubin 1974).

Attempting to observe both potential outcomes in the same individual is only possible under strict assumptions that are generally considered untenable. For example one

could observe the woman's health status before and then after treatment but the role of temporality is generally non-negligible, resulting in a comparison that is not causal in its nature Rubin (1974). Per Rubin (1977), "one cannot simply look at the plot of Posttest on Pretest and properly estimate treatment effects." It is also generally inadvisable, unless rigorous safeguards are enacted, to compare the outcome of a treated individual to another untreated individual, since any two individuals are likely to be different or have different pretreatment variables that are important to consider. A good estimate of the causal effect of a treatment can result from comparing statistical summary of the health outcomes of the set of individuals who were treated to that of the set of individuals who were untreated, provided that these two groups of individuals had similar pretreatment variables.

Given a sample of participants in a well-defined and well-executed clinical trial, if each individual is randomized to receive treatment or to not receive treatment, the difference in the average health outcomes in the two groups of patients leads to an unbiased estimate of the causal effect of the treatment on the outcome of interest. Randomized trials have been the gold standard for making causal conclusions (Schulz and Grimes 2002) because when they are properly designed and executed, none of the factors that influence an individual's potential outcome can affect that individual's treatment status as a result of the randomization of individuals to treatment. A factor influences both an individual's potential outcomes and the likelihood that the individual obtains treatment is often called a confounding variable (Rosenbaum and Rubin 1983b), or a confounder (Greenland and Robins 1986, Greenland et al. 1999b). Failure to control for confounders typically results in biased estimates of causal effects. Randomized trials can eliminate confounding by design, but suffer the drawbacks that they are usually costly to implement, are carried out on a relatively small portion of the population of interest (Grimes and Schulz 2002). Additionally, randomized trials can be infeasible or

unethical in certain situations.

In these cases, causal inference should be addressed from nonrandomized studies. Nonrandomized studies are relatively less expensive and can follow a larger share of the population of interest compared to randomized studies. Also called nonexperimental (LaLonde 1986, Dehejia and Wahba 1999; 2002) or observational (Cochran and Chambers 1965) studies, these study designs exhibit the potential for confounding. Traditional statistical techniques often do not adequately control for confounding and can provide biased estimates of causal effects, among other undesirable results (Holland 1986).

Rubin (1974; 1977; 1978; 1980) pioneered the use of counterfactuals and potential outcomes to develop a set of methods that can eliminate confounding in observational studies given certain assumptions. Rubin (1974) describes a process of matching treated to untreated individuals such that one untreated individual and one treated individual are matched "prior to the initiation of treatments on all variables thought to be important in the sense that they causally affect" outcome. Rubin (1974) argues that even absent randomization, "having closely 'matched' trials increases the closeness of the calculated experimental minus control difference." Rubin (1977) discusses methods to control for these important variables by blocking.

Rubin (1978) expanded on this definition of these important pretreatment variables, defining "ignorable" treatment assignment when each individual's potential outcomes were statistically independent of the individual's treatment, controlling for a sufficient set of the individual's covariates. In the notation of Dawid (1979),

$$Y_i(a) \perp A_i \mid L_i \text{ for all units } i \text{ and for } a = 0, 1, \tag{1.3}$$

where L_i is a sufficient set of covariates. Similar concepts and related phrases strong

ignorability (Rosenbaum and Rubin 1983b) exchangeability or conditional exchangeability (Greenland and Robins 1986), and the assumption of no unmeasured confounders (Robins et al. 2000).

1.3.2 Matching and Inverse Probability Weighting

Controlling for confounding variables takes many forms, but the main goal is to improve inference by reducing bias in the estimates. Two of the many popular methods that can be used for causal are matching and inverse probability of treatment weighting (IPW). Each of these can be carried out nonparametrically or with parametric modeling assumptions. Here we present a short review of matching and IPW in nonparametric settings when the dimensionality of these important covariates is low. Later we review how to carry out analysis when the dimension of the covariates is high in Section 1.3.4 by introducing the propensity score and discussing modeling assumptions.

Matching methods take many types of forms. When the dimensionality of covariates is low, matches can be defined nonparametrically on exactly identical combinations of covariates, or using a Mahalanobis distance metric between individuals, among others (Rubin 1974, Imai et al. 2008, Stuart 2010, King and Nielsen 2016). Matching has gained popularity because it is highly interpretable and easy to implement, and can require little more than checks to ensure covariate balance. More modern matching techniques include methods to programmatically determine matches in a way that can relieve the necessity for the investigator to do these checks (Iacus et al. 2012).

Inverse probability weighting by the conditional probability of treatment, often called IPTW or simply IPW, relies on (1.4) to create an balanced "psuedopopulation" (Robins et al. 2000, Cole and Hernán 2004) in which differences between the

two treatment groups result in unbiased estimates of causal effects. IPW is most often implemented using parametric assumptions and modeling the probability of treatment conditional on covariates, but nonparametric IPW estimation is also possible and asymptotically efficient (Huber et al. 2013). Rosenbaum (1987) described the connection between IPW and literature on survey sampling weights. An advantage of IPW is that, unlike matching, it does not necessarily require individuals in each treatment group to have the same covariates. However, it can suffer from poor performance when the probability weights are large.

1.3.3 Estimating Causal Effects from Observational Studies

Many studies have shown that insufficiently controlling for confounders and failing to satisfy the conditional exchangeability assumption (1.3) results in biased estimates of causal parameters (Rosenbaum and Rubin 1983a). The ability of researchers to appropriately identify the correct set of potential confounders necessary to condition upon is a major hurdle to the field of causal inference from observational data (Robins and Greenland 1986, Rubin and Thomas 1996, Brookhart et al. 2006, Austin 2011, Westreich et al. 2011, Schisterman et al. 2009). This assumption is said to be untestable as its validity cannot be examined from observable data (Holland 1986). Causal diagrams are important tools for the researcher interested in determining conditional exchangeability (Pearl 1995, Greenland et al. 1999a, Ogburn et al. 2014). For example, Richardson and Robins (2013) introduced Single-World Intervention Graphs as a way to draw causal diagrams that explicitly show potential outcomes. Many methods have been proposed as a way to mitigate the risk of violating the assumption of conditional exchangeability (Rubin 1997, Schneeweiss et al. 2009, VanderWeele and Shpitser 2011). In the sequel, it is assumed that L_i satisfies conditional exchangeability (1.3).

1.3.4 The Propensity Score

When dimensionality of potential confounders is high, either due to a large number of covariates or the presence of continuous covariates, nonparametric techniques may no longer be appropriate. In this case, models of the distribution of treatment conditional on covariates that rely on parametric assumptions are often used in conjunction with matching and IPW. To this end, a very popular concept is the conditional probability that an individual obtained the treatment that it was observed to have experienced, or the propensity score. Rosenbaum and Rubin (1983b) introduced the propensity score $e(L_i) = \Pr(A_i = 1 | L_i = l_i)$ as the coarsest (scalar) function of the potential confounders that creates balance between the two treatment groups. They show that by balancing the two groups, the propensity score also removes bias from confounding as in (1.3):

$$Y_i(a) \perp A_i \mid e(L_i) \text{ for all } i \text{ and for } a = 0, 1 \tag{1.4}$$

The propensity score also provides a clear conceptual link between drawing causal inferences from randomized experiments and doing so from nonrandomized studies. In experiments, a study unit is randomized to treatment in a way that generalizes to a coin flip or the roll of a weighted die. For a study unit with covariates L_i in an nonrandomized study, and assuming strong ignorability on L_i (1.4), the unit's propensity score $e(L_i)$ is the true probability that the unit obtains treatment. That is, the study unit obtains treatment by the hypothetical process of rolling of a weighted die or flipping of a weighted coin where the weight is equal to $e(L_i)$. Using the propensity score to link observational studies that lack randomization to some form of similar or "ideal" randomized experiment allows investigators to expand the role of causal inference in nonexperimental studies (Rubin 1974, Rosenbaum 2002).

A great deal of research estimates causal quantities by matching individuals from

different treatment groups when they have similar values of their estimated propensity score (Rosenbaum and Rubin 1983b, Rubin and Thomas 1996, Dehejia and Wahba 1999, Rubin and Thomas 1996, Dehejia and Wahba 2002, Smith and Todd 2005, Ho et al. 2007, Stuart 2010, Abadie and Imbens 2016). Similarly, IPW methods have been used in many scenarios when the weights are derived from estimated propensity scores (Rosenbaum 1987, Robins et al. 2000, Robins and Finkelstein 2000, Hirano et al. 2003, Cole and Hernán 2004). There has been some debate as to which of these two standard methods outperforms the other. For example, Frölich (2004a) found that matching performed better than IPW, yet Busso et al. (2014) found that IPW performed better than matching, and Huber et al. (2013) shows that each method has its advantages and its disadvantages. Both methods have been used for important contributions to the causal literature, and yet show great potential for future adaptations.

1.3.5 Estimating the Propensity Score

In observational studies, the propensity scores are unknown; they must be estimated from observed data. Even when (1.3) is satisfied and the investigator has chosen an appropriate combination of covariates to control for, there is still the risk for bias from inappropriate modeling assumptions. Fitting parametric and some semi-parametric models to estimate the propensity score is also subject to the assumption of correct model form. That is, even when (1.3) is satisfied, any model specified by the researcher must have the functional form of the true underlying distribution of treatment, e.g. including interactions and higher order terms as appropriate (Rosenbaum and Rubin 1984, Drake 1993, Austin 2011, Vansteelandt et al. 2012). Failure to satisfy this assumption results in biased estimates of causal effects. Like (1.3), this assumption is untestable. Research has been carried out at length to determine sensitivity of various model fitting and variable selection techniques to this assumption (Brookhart et al.

2006, Setoguchi et al. 2008, Westreich et al. 2010, Lee et al. 2010, Watkins et al. 2013, Wyss et al. 2014). Recently, various techniques for estimating propensity scores arising from the machine learning field have been proposed as a robust alternative to traditional, fully parametric models (Woo et al. 2008, McCaffrey et al. 2013, Zhu et al. 2015, Pirracchio et al. 2015, Pirracchio and Carone 2016).

1.4 Multiple Forms of Treatment

Investigators often assume SUTVA in order to simplify analyses, but this assumption is not always appropriate. Notable examples of when SUTVA is an untenable assumption include studies with three or more distinct levels of treatments (Lechner 2001, Frölich 2004b, Zanutto et al. 2005, Spreeuwenberg et al. 2010, Cadarette et al. 2010, Cattaneo 2010, Feng et al. 2012, McCaffrey et al. 2013, Rassen et al. 2013, Fong and Imai 2014, Linden and Yarnold 2016, Yang et al. 2016) or a continuous or dose-response relationship (Robins et al. 2000, Imbens 2000, Foster 2003, Hirano and Imbens 2004, Imai and Van Dyk 2004, Kluve et al. 2012, Egger and Von Ehrlich 2013, Fong and Imai 2014, Kreif et al. 2015, Schuler et al. 2016).

Methods appropriate for multilevel or continuous treatments are often based on the works by Robins et al. (2000), Imbens (2000), Imai and Van Dyk (2004), Hirano and Imbens (2004). In particular, Yang et al. (2016) recently introduced a matching method based on principles put forth by Imbens (2000) that shows great potential for future adaptations due to two small adjustments from most previous methods. Firstly, Yang et al. (2016) assumes a "weak" version of (1.3) that does not require individuals to be good matches across all levels of treatment, instead requiring only that they are good matches in the two levels of treatment corresponding to their observed treatment. Also, while many matching methods directly estimate causal effects by using matched pairs to estimate contrasts, Yang et al. (2014) instead uses the matched pairs to impute

all potential outcomes for each individual, thereby estimating full information of all counterfactual scenarios in some cases.

Allowing for multiple treatments still runs the risk of violating other assumptions. For example, any treatment model must still have the correct form, which is perhaps even more challenging with multiple treatments than when assuming SUTVA. Recently, authors have presented flexible modeling techniques drawing from machine learning to mitigate the risk of incorrectly specifying propensity scores when the treatment is multilevel or continuous (McCaffrey et al. 2013, Kreif et al. 2015, Zhu et al. 2015).

1.5 Interference

There are major conceptual and mathematical differences in the estimands that allow for interference compared to those that consider non-binary individualized treatments but do not allow for interference. A challenge is that when interference is present, each individual may be subjected to what seems like a binary and individualized intervention, but that individual may have many more than two potential outcomes arising from the interventions that other individuals are subjected to.

This document concerns itself with interference, which appears in the literature as far back as Cox (1958) and Neyman (1935). Rubin (1990) noted the potential for interference arising from interpersonal interactions, in his case in the study of educational interventions. Perhaps an instructive example of interference is in crossover or changeover trials, when an individual's outcome in the second stage may be influenced not only by the treatment they receive in that period but also the treatment they had received in a different period (Grizzle 1965). For many years, interference was primarily considered in the language of a "technical error" (Neyman 1935, Rubin 1980), a "major issue" Rubin (1990), or something in need of "washout" (Verhave et al. 1959, Brown Jr 1980, Koch et al. 1980). Halloran and Struchiner (1995) argue for the

careful consideration of the effects related to interference, saying, "The interference in agricultural experiments, for example, is a nuisance that we try to be rid of . . . however . . . the exposure to infection provided by the other members of the population in infectious diseases, either directly or via vectors, is essential to transmission as well as for evaluating the effects of the intervention."

After Rubin (1978; 1990), Halloran and Struchiner (1995) were among the first to consider interference in the potential outcomes framework. They reintroduce in the language of potential outcomes the direct, indirect, total and overall effects of treatment in the presence of interference from Halloran et al. (1991). Hong and Raudenbush (2006) investigated peer effects in educational classrooms, desiring to determine whether removing or adding a child to the classroom changed the learning outcomes of the other children who were in the classroom. In this study, most of the students did not change classrooms (i.e. were not subject to a change in individualized intervention status) but the composition of their classrooms was changed by the addition or subtraction of other students, providing an interesting example of how interference can arise from social interactions. Sobel (2006) introduced the term "partial interference" for describing the assumption that interference was possible within distinct groups or clusters, but no interference is assumed between clusters, calling these neighborhood effects or spillover effects. Hudgens and Halloran (2008) proposed estimands for the effects of interference that are estimable using a two-stage cluster-randomized trial.

Other areas in which causal effects in the presence of interference have been studied include criminology (Verbitsky-Savitz and Raudenbush 2012), spatial analyses (Zigler et al. 2012, Graham et al. 2013), medical imaging (Luo et al. 2012), econometrics (Sobel 2006, Manski 2013, Arpino and Mattei 2016), education (Basse and Feller 2016, Kang and Imbens 2016), political science (Bowers et al. 2013), public policy (Graham 2011, Baird et al. 2016), sociology (Gangl 2010, Aronow 2012), and social media and

network analysis (Ugander et al. 2013, VanderWeele and An 2013, Toulis and Kao 2013, Kramer et al. 2014, Ogburn et al. 2014, van der Laan 2014, Kim et al. 2015, Eckles et al. 2016, Sofrygin and van der Laan 2017, Athey et al. 2017). Clearly, the study of causal inference in the presence of interference is very popular at this time.

Concerns for estimating the direct effect of vaccines go back to at least Greenwood and Yule (1915), Ross (1916). Halloran and Struchiner (1995) were the first to use the potential outcomes framework to discuss the effects of partial interference in studies of infectious disease. Hudgens and Halloran (2008) proposed a set of estimands that are estimable in a two-stage randomized trial. VanderWeele and Tchetgen Tchetgen (2011b) proposed estimands for a similar yet two-stage trial. Tchetgen Tchetgen and VanderWeele (2012) introduced estimators for those estimands, both for the two-stage trial and also for use in observational studies using inverse probability weights. Perez-Heydrich et al. (2014) derived closed-form asymptotic variance estimators for the estimators when modeling the propensity score. Liu et al. (2016) introduced a new set of estimators without the need to assume partial interference, with corresponding variance estimators when assuming partial interference. Related work has been carried out by Liu and Hudgens (2014), Rigdon and Hudgens (2015), VanderWeele and Tchetgen Tchetgen (2011a), VanderWeele et al. (2014), Forastiere et al. (2016b;a), Carnegie et al. (2016), Manski (2016), along with software for estimating causal effects in the presence of interference introduced in Saul (2017), Rigdon (2015), Barkley (2018).

1.6 Summary and Proposed Research

Interference is an important concern in the design and analysis of both randomized and observational studies, and should be accounted for. Randomized trials can be cleverly designed to disentangle these effects to at least a certain extent. Correctly accounting for interference is more challenging in a nonrandomized study, and additional

care is required for unbiased estimation and causal inferences. One major goal that is addressed in the following two chapters is proposing new estimators for existing estimands. Another main concern is defining new estimands that can connect observational data to questions of scientific interest, which is undertaken in the final chapter.

In Chapter 1.6, we introduce a method for more robust estimation of propensity scores in the presence of partial interference. We assume a particular correlation structure, and propose an IPW estimator that uses machine learning as a form of model selection for the fixed effect predictors. The finite-sample performance of the estimators is investigated in a simulation study, and the proposed methods are illustrated in a data analysis of the Matlab, Bangladesh cholera vaccine study.

A matching method for estimating causal effects in the presence of partial interference is introduced in Chapter 2.8. Matching methods show promise for exhibiting less bias than the existing IPW estimators in a variety of scenarios due to the relative instability of the IPW. Asymptotic properties are demonstrated, and an estimator of the asymptotic variance is proposed. The finite-sample performance of this estimator is investigated in a simulation study, and compared to the performance of existing IPW estimators. The proposed methods are illustrated in an analysis of the Matlab, Bangladesh cholera vaccine study.

In Chapter 3.10, we propose new estimands tailored for use with observational studies under the assumption of partial interference. These estimands describe counterfactual scenarios in which treatment may be correlated, and so they may be more relevant to public health researchers. Identifiability of the estimands from nonexperimental data is be discussed, and estimators are proposed. The estimators are shown to be consistent and asymptotically normal, and their finite-sample performance is evaluated in a simulation study. The proposed methods are illustrated in an application to a cholera vaccine study in Matlab, Bangladesh.

CHAPTER 2: MACHINE LEARNING FOR ESTIMATING CLUSTER PROPENSITY SCORES IN PARTIAL INTERFERENCE

2.1 Introduction

Inferring causal effects from an observational study is challenging because participants are not randomized to treatment. Using traditional associative statistical methods in an attempt to draw causal conclusions from observational studies is often subject to confounding and bias. A tool for inference is to model the distribution of treatment on predictors in a sample of data; this is often used when carrying out estimation via inverse probability-weighting (IPW) by the estimated propensity score.

Performance of these methods are subject to the investigator's choice of treatment models, a decision whose appropriateness cannot be fully determined from observable data (Robins and Greenland 1986, Drake 1993, Brookhart et al. 2006, Vansteelandt et al. 2012). This topic has been studied at some length in the case where there are exactly two versions of treatment, and some authors have proposed methods for the case where there are more than two versions of treatment (McCaffrey et al. 2013, Kreif et al. 2015, Zhu et al. 2015). However, these decisions and their corresponding consequences have not been explored thoroughly in the case where an individual's treatment may affect another individual's outcome, and "interference" (Cox 1958) is said to be present. Interference presents analytical challenges in infectious disease epidemiology among many other fields of research. The focus of this paper is to introduce statistical methods that reduce the risk of bias due to model mis-specification in observational studies in the presence of interference.

A major challenge to drawing inference when interference is present is the proliferation of the number of potential treatments for an individual. Furthermore, non-negligible correlation is likely to exist between individual treatments, presenting additional challenges for treatment modeling in this scenario. Many existing methods for estimating causal effects from observational studies in the presence of interference rely on fitting fully parametric logistic mixed effects models (see e.g., Perez-Heydrich et al. (2014), Liu et al. (2016) and Barkley et al. (2017)), which are at risk of bias due to model mis-specification. There is currently a dearth of flexible statistical techniques that can be used to model correlated treatment for estimating propensity scores in the presence of interference (Liu et al. 2016). We propose using the GMERT algorithm (Hajjem et al. 2017) as a data-adaptive model selection method for estimating propensity scores in this scenario, a method we call GMERT-IPW.

The remainder of this paper is as follows. In Section 2.2 we provide a brief overview of the potential outcomes framework and treatment modeling assumptions in causal inference. In Section 2.3 we review interference and the causal estimands of interest for this study. Section 2.4 reviews existing IPW estimators that rely on fully parametric logistic mixed models, and challenges for treatment modeling in this scenario. The proposed GMERT-IPW method is introduced in Section 2.5. Finite-sample performance of GMERT-IPW is investigated in simulations in Section 2.6. The proposed method is illustrated in an analysis of a large cholera vaccine study in Matlab, Bangladesh in Section 2.7. This paper concludes with a discussion.

2.2 Treatment Modeling Assumptions in Causal Inference

The main assumption necessary for drawing casual effects of a treatment intervention A on a subsequent outcome Y from observed data is that each individual's potential outcomes $Y(a)$ are unaffected by observed treatment conditional on a sufficient set of

observed pretreatment variables, L. This assumption of conditional exchangeability can be written using the notation of Dawid (1979) as

$$Y_i(a) \perp A_i \mid L_i \text{ for all units } i \text{ and for } a = 0, 1. \tag{2.5}$$

Investigators often control for the probability that an individual obtains treatment conditional on the sufficient set of confounders L called the propensity score (Rosenbaum and Rubin 1983b), which can be written as

$$\Pr(A_i|L_i) \text{ for all } i \tag{2.6}$$

Rosenbaum and Rubin (1983b) showed that when L satisfies (2.5), then conditioning on the scalar propensity score balances a sample of of observed data such that conditional exchangeability is satisfied:

$$Y_i(a) \perp A_i \mid \Pr(A_i|L_i) \text{ for all } i \text{ and for } a = 0, 1. \tag{2.7}$$

IPW estimators often rely on propensity scores, as do various other methods for drawing inference from non-experimental data including some matching estimators.

Propensity scores are unknown in observational studies and must be estimated from observed data; in general, the propensity score for each unit i in the sample is estimated after fitting some sort of statistical model for treatment A on covariates L. For example, an investigator must make an untestable assumptions that a set of confounders L is sufficient to satisfy (2.5). More relevant to this paper is the untestable assumption that the investigator's chosen statistical model for treatment on confounders has the appropriate form (Drake 1993, Vansteelandt et al. 2012). Estimates of causal effects may be biased if either of these assumptions are inappropriate.

Here and in the sequel, we assume that (2.5) is satisfied to focus on modeling assumptions. More explicitly, we say that the assumption of correct model form is violated when the following are simultaneously true: (2.5) (and thus (2.7)) is satisfied, yet the data sample remains unbalanced - even after adjustment with estimated propensity scores - as a result of inappropriate modeling assumptions. We refer to these as violations of the functional form assumptions (FFA), and they can result in biased estimates of causal quantities, among other undesirable inferential properties.

Studies have been carried out to determine sensitivity to the FFA of various model fitting and variable selection techniques (Brookhart et al. 2006, Setoguchi et al. 2008, Westreich et al. 2010, Lee et al. 2010, Watkins et al. 2013, Wyss et al. 2014). Recently, new techniques for estimating propensity scores arising from the field of machine learning have been proposed as a robust alternative to fully parametric models for scenarios in which treatment is binary (Woo et al. 2008, Pirracchio et al. 2015, Pirracchio and Carone 2016), multilevel (McCaffrey et al. 2013), or continuous (Kreif et al. 2015, Zhu et al. 2015, Linden and Yarnold 2016). However, sensitivity to the FFA has received little attention when allowing for interference, and there is a lack of flexible models for estimating propensity scores in this scenario (Liu et al. 2016).

This manuscript is concerned with developing a modeling strategy more robust to violations of the FFA in the presence of partial interference. Existing IPW estimators in this setting rely on a logistic mixed model with a random intercept for cluster or group membership to account for some amount of presumed treatment correlation, as discussed in Section 2.4.1. These models are at risk for model misspecification and violations of the FFA due to their fully parametric specification; modeling challenges in this scenario are discussed in Section 2.4.2. We propose to use the GMERT algorithm introduced in Hajjem et al. (2017), which draws from the field of machine learning and also allows for estimation of variance components as in traditional mixed models.

The proposed GMERT-IPW methods are introduced in Section 2.5, and they rely on sample-splitting and also the existing IPW estimators discussed in Section 2.4.

2.3 Estimands for Partial Interference

Interference is when one individual's outcome may be affected by another individual's treatment, and it considered a violation or relaxation of SUTVA (Rubin 1980, VanderWeele 2009). It may be reasonable to assume that individuals can be partitioned into clusters such that interference is plausible within any cluster, but not between individuals in two distinct clusters. This assumption has been termed as "partial interference" (Sobel 2006) or "clustered interference" (Barkley et al. 2017). Partial interference is assumed throughout.

This paper considers the estimands described in Tchetgen Tchetgen and VanderWeele (2012) for use in observational studies. Like those in Hudgens and Halloran (2008), these estimands arise from a two-stage cluster-randomized trial. Consider a super-population of clusters of individuals; for consistency in notation, index the clusters by i. Let any cluster i be comprised of $j = 1, \ldots, N_i$ individuals. Let Y_{ij} denote the observed outcome for individual j in cluster i where e.g., $Y_{ij} = 1$ perhaps indicates infection and equals 0 otherwise. Let $Y_i = (Y_{i1}, \ldots, Y_{iN_i})$ be the vector of outcome statuses for the members of the cluster. Let A_{ij} denote the treatment status for individual j in cluster i, where $A_{ij} = 1$ if the individual was treated and 0 otherwise, and let $A_i = (A_{i1}, \ldots, A_{iN_i})$ be the vector of treatment statuses for the cluster. Let $a \in \mathcal{A}(N_i)$ be a binary vector of length N_i, where $a = (a_1, \ldots, a_{N_i})$. Let $a_j \in \{0, 1\}$ denote the j^{th} element of a, and let $a_{-j} = (a_1, \ldots, a_{j-1}, a_{j+1}, \ldots, a_n)$ denote the $(n-1)$-dimensional subvector of a excluding a_j. Define $Y_{ij}(a)$ to be the potential outcome for individual j in cluster i when cluster i experiences treatment $a \in \mathcal{A}(N_i)$.

Tchetgen Tchetgen and VanderWeele (2012) describes a two-stage cluster-randomized

trial where individuals are administered treatment according to a "type B parameterisation." Let α and α' be two treatment allocation strategies that change the distribution of treatment. In such a trial, clusters are assigned to either the α or α' arm, and when a cluster is assigned to the α arm then the individuals in that cluster are randomized to treatment with equal probability which we denote $\alpha \in [0, 1]$. We represent the distribution of treatment in clusters in the α arm by the Bernoulli-type product

$$\pi(a, \alpha) = \prod_{j=1}^{N} \alpha^{a_j}(1-\alpha)^{(1-a_j)}. \tag{2.8}$$

The estimands of interest are described by averages of potential outcomes under an allocation strategy. For example, when cluster i is assigned to the α arm, then it experiences potential outcome $Y_i(a)$ with probability $\pi(a, \alpha)$. An average of its potential outcomes under α is then

$$\frac{1}{N_i} \sum_{j=1}^{N_i} \sum_{a \in \mathcal{A}(N_i)} Y_{ij}(a)\pi(a, \alpha).$$

Define the population mean potential outcome to be

$$\mu(\alpha) = \mathbb{E}\left[\frac{1}{N_i} \sum_{j=1}^{N_i} \sum_{a \in \mathcal{A}(N_i)} Y_{ij}(a)\pi(a, \alpha)\right]. \tag{2.9}$$

One type of causal effects are the Overall Effects, defined to be differences in the population mean potential outcomes from two different allocation strategies:

$$\mathrm{OE}(\alpha, \alpha') = \mu(\alpha) - \mu(\alpha'). \tag{2.10}$$

Similarly, for $z \in \{0, 1\}$, define:

$$\mu(z, \alpha) = \mathbb{E}\left[\frac{1}{N_i}\sum_{j=1}^{N_i}\sum_{a' \in \mathcal{A}(N_i - 1)} Y_{ij}(z, a')\pi(a', \alpha)\right], \tag{2.11}$$

where $\pi(a_{-j}, \alpha) = \pi(a, \alpha)/\pi(a_j, \alpha)$. The Direct, Indirect, and Total Effects are parameterized here as:

$$\text{DE}(\alpha) = \mu(1, \alpha) - \mu(0, \alpha) \tag{2.12}$$

$$\text{IE}(\alpha, \alpha') = \mu(0, \alpha) - \mu(0, \alpha') \tag{2.13}$$

$$\text{TE}(\alpha, \alpha') = \mu(1, \alpha) - \mu(0, \alpha') \tag{2.14}$$

Similar estimands are defined in Liu et al. (2016); in the case of partial interference when each cluster is the same size, the estimands are identical to those presented above.

2.4 Existing IPW Estimators

Consider a sample of $i = 1, \ldots, M$ clusters, where each cluster has N_i individuals and an i.i.d. copy of the random variables (L_i, A_i, Y_i). As above, Y_i is a N_i-vector of observed outcomes and A_i is a binary N_i-vector of observed treatments. Here, $L_i = (L_{i1}, \ldots, L_{iN_i})$ is a $(N_i \times p)$-dimensional matrix of covariates for the cluster, where p equals the number of pre-treatment covariates. The pre-treatment covariates for individual j in the cluster are represented by the $(1 \times p)$-dimensional row vector $L_{ij} = (L_{1ij}, \ldots, L_{ijp})$. The following IPW estimators for the target estimands assume that L_i satisfies conditional exchangeability and positivity (i.e, $\Pr(A_i = a|L_i) > 0$ for all i and all $a \in \mathcal{A}(N_i)$):

$$\widehat{\mu}(\alpha) = \frac{1}{M}\sum_{i=1}^{M}\frac{1}{N_i}\sum_{j=1}^{N_i}\frac{Y_{ij}\pi(A_i, \alpha)}{\Pr(A_i|L_i)} \tag{2.15}$$

and for $z = 0, 1$,

$$\widehat{\mu}(z, \alpha) = \frac{1}{M} \sum_{i=1}^{M} \frac{1}{N_i} \sum_{j=1}^{N_i} \frac{Y_{ij} I(A_{ij} = z) \pi(A_{i,-j}, \alpha)}{\Pr(A_i | L_i)} \tag{2.16}$$

where $\Pr(A_i|L_i)$ denotes the propensity score for group or cluster i. Discussion of this propensity score is presented in Section 2.4.1 below. The four types of target causal effects can be estimated by taking appropriate differences of the above estimators. See e.g., Tchetgen Tchetgen and VanderWeele (2012, Section 5.2) or Perez-Heydrich et al. (2014) for further discussion of these estimators.

Liu et al. (2016) proposed "stabilized" IPW estimators for similar estimands:

$$\hat{\mu}_{Hajek}(\alpha) = \hat{N}^{-1} \sum_{i=1}^{N} \frac{\bar{Y}_i \pi(A_i, \alpha)}{\Pr(A_i = a | L_i = l_i)} \tag{2.17}$$

where now $\hat{N}$ is replaced with an estimated term. Liu et al. (2016) proposes two methods for this term; for example the second method is:

$$\hat{N} = \sum_{i=1}^{N} \frac{\pi(A_i, \alpha)}{\Pr(A_i = a | L_i = l_i)}.$$

In the special case when all clusters have the same number of units, the estimands that these estimators target are identical to the estimands from Tchetgen Tchetgen and VanderWeele (2012). Each of these IPW estimators is consistent and asymptotically Normal when the treatment model is correctly specified. The asymptotic variance can be estimated with a sandwich variance estimator using standard estimating equation theory; see e.g., Stefanski and Boos (2002) or Saul and Hudgens (2017).

2.4.1 Fully Parametric Logistic Mixed Model for Treatment

In this section we provide further details on the propensity scores for the IPW estimators described above. As in Barkley et al. (2017), we refer to

$$\Pr(A_i = a_i|L_i) = \Pr(A_{i1} = a_{i1}, A_{i2} = a_{i2}, \ldots, A_{iN_i} = a_{iN_i}|L_i) \tag{2.18}$$

as the cluster propensity score to emphasize that it is the joint probability of multiple individual treatment statuses.

When treatment is uncorrelated then a logistic GLM or machine learning classification techniques may be appropriate to model the relationship between treatment and pre-treatment covariates and estimate (2.18). However, when interference is plausible then a reasonable assumption is that treatment is correlated. As in Perez-Heydrich et al. (2014) and Liu et al. (2016), we assume that the correlation structure can be described by a random intercept for each cluster. We assume that the cluster propensity score has the form

$$\Pr(A_i = a_i|L_i) = \int \prod_{j=1}^{N_i} p_{ij}^{a_{ij}} (1 - p_{ij})^{(1-a_{ij})} d\Phi(b|\sigma) \tag{2.19}$$

where $p_{ij} = \Pr(A_{ij} = 1|L_{ij}; b_i)$ represents the probability that individual j is exposed to treatment conditional on L_{ij} and also on the cluster's random intercept b_i. That is, for some function f of the pre-treatment covariates, we assume

$$p_{ij} = \mathcal{L}^{-1}\left(f(L_{ij}) + b_i\right), \tag{2.20}$$

where the random intercept is assumed to follow a Gaussian distribution with mean zero: $b_i \sim N(0, \sigma)$.

Existing methods assume a fully parametric logistic mixed model such that

$$f(L_{ij}) = L_{ij}^{\intercal}\beta. \tag{2.21}$$

For simplicity, we refer to a logistic mixed effects model (with a random intercept for cluster) in the sequel as an "LMM." In this fully parametric method, the parameters β and σ would be estimated from a model fit with maximum likelihood techniques, such as adaptive Gaussian Quadrature with the `glmer` function in the `lme4` package (Bates et al. 2015). Then, these parameters would be used to estimate $\Pr(A_i = a_i|L_i; \hat{\beta}, \hat{\sigma})$ for each cluster i, which would then be used to estimate and draw inference about the target casual estimands.

2.4.2 Modeling Challenges for Estimating Cluster Propensity Scores

The performance of the LMM model to estimate cluster propensity scores will depend on, among other things, whether the LMM model has the correct functional form for the covariates in the model. In this work, we assume that (2.19) and (2.20) are correct. We focus on a method for estimating the cluster propensity score that uses an alternatives to the linearity assumption in (2.21).

When no interference is assumed (e.g., SUTVA), machine learning methods have been shown to be more robust to the FFA. However, when interference is present, it's likely that treatment is also correlated. There is a growing amount of research that has investigated methods to estimate propensity scores assuming no interference when individuals' treatment exposures may be correlated, e.g., see Arpino and Mealli (2011), Cannas et al. (2012), Li et al. (2013), Arpino and Cannas (2016), Schuler et al. (2016) or Yang (2017). However, when treatment is correlated and clustered interference is assumed, then the interest is no longer in (individual) propensity scores, but rather cluster propensity scores. The interest is to integrate over the (presumed) distribution

of the random effects as shown in (2.19) (rather than, say, predicting with empirical BLUPs). This may present additional methodological challenges. Here we consider the application of a recently-proposed GMERT algorithm (Hajjem et al. 2017). Applying GMERT to a training sample allows us to make assumptions about the functional form in a data-adaptive manner.

2.5 Proposed GMERT-IPW Method

Hajjem et al. (2017) recently introduced a method combining combining decision trees and mixed modeling techniques, called Generalized Mixed Effects Regression Trees or GMERT; it is applicable with correlated binary responses. We present some of the motivating ideas behind the GMERT algorithm in Section 2.5.1, and we refer the interested reader to Hajjem et al. (2017) for further details. Then, in Section 2.5.2 the GMERT-IPW method for estimating the target estimands is proposed.

2.5.1 GMERT

Hajjem et al. (2017) propose to model correlated binary response data with a logit-link function on the probability of treatment, where the probability is assumed to have an underlying, latent random variable with respect to the cluster identifier (i.e., a random intercept). In a traditional LMM, the form of the fixed effects is pre-specified, and the model simply estimates the fixed effects. GMERT uses a decision tree algorithm (i.e., `rpart` (Therneau et al. 1997)) to simultaneously determine the form of the fixed effects while also obtaining predicted values for them. For our purposes with a split sample, the structure will be more important, although the predicted values can be used for cross-validation as we do in Section 2.7. The GMERT algorithm extends the MERT algorithm in Hajjem et al. (2011) from continuous responses to binary responses by using penalized quasi-likelihood (PQL).

Traditionally, PQL is used to fit a weighted linear mixed effects model in an iterative fashion to obtain approximate parameter estimates for generalized mixed effects models (Breslow and Clayton 1993). The PQL algorithm uses a Taylor series linearization of the response to create what is called a "working response," which can take on values on the real number line. Many frequentist mixed model estimation procedures transform the likelihood (see e.g., Fitzmaurice et al. (2012)); PQL is key here because it transforms the data as well. In the GMERT algorithm, the working response is "de-correlated" by subtracting empirical BLUPs to create an "adjusted working response" that can be used to with supervised machine learning techniques, e.g., weighted decision trees. A sketch of the GMERT algorithm is provided in Section A.2.

2.5.2 GMERT-IPW

We propose to use the structure of the predictors from the tree algorithm recovered by GMERT as a means of model selection. That is, we ignore the estimated parameters on the training set, and instead fit a LMM on the testing set of data with no intercept but one parameter for each of the regions as designated by the tree; we refer to this modeling technique as a GMERT-LMM. The GMERT-IPW method is to use a GMERT-LMM to estimate cluster propensity scores for IPW estimation of target estimands in the presence of partial interference. We present the proposed method in more detail below.

Label $\mathcal{S}$ the entire sample of all $i = 1, \ldots, M$ clusters, and partition this into a training sample $\mathcal{S}_{train}$ of M_{train} clusters, and a testing sample $\mathcal{S}_{test}$ of the remaining $M_{test} = M - M_{train}$ clusters, where each cluster belongs entirely to either the training or the testing sample. We apply GMERT to $\mathcal{S}_{train}$, and recover the decision tree. The decision tree describes a partition $\Theta = \{r_1, \ldots, r_T\}$ of T-many regions of the space of all possible covariates. These regions are distinct ($r_t \cap r_{t'} = \emptyset$ when $t \neq t'$) and they span

the predictor space ($\bigcup_{t \leq T} r_t = \mathbb{L}$). These regions describe a set of rules to partition any sample of data that shares the same predictor space, e.g., $\mathcal{S}_{test}$.

We can use the partition as a rule to transform the covariates of all individuals in all clusters in $\mathcal{S}_{test}$ into the T regions (note that the predicted values and eBLUPs from the training set are not used). That is, for all $j \leq N_i$ in all clusters $i \in \mathcal{S}_{test}$, we can transform $X_{ij} = \Theta(L_{ij})$, where X_{ij} is a $(1 \times T)$-dimensional row vector of indicator variables corresponding to the region in which L_{ij} falls.

We fit an LMM on $\mathcal{S}_{test}$ for treatment A on the transformed predictors $X = \Theta(L)$. This model can be described with cell-means coding, and so maximum likelihood estimation can be carried out with existing software (i.e., adaptive Gaussian Quadrature using `glmer` (Bates et al. 2015)). Thus the use of the GMERT algorithm is perhaps aptly described as a model (and variable) selection procedure to inform the LMM. For the GMERT-LMM we assume (2.19) and (2.20), yet instead of (2.21) we assume:

$$f_\Theta(L_{ij}) = X_{ij}^\intercal \beta_\Theta. \tag{2.22}$$

The GMERT-IPW method is completed by using the GMERT-LMM to estimate cluster propensity scores and to target existing estimands in Section 2.3 with the existing IPW estimators in Section 2.4.

This procedure has the potential to reduce bias due to model misspecification, which has not yet been studied in this scenario. Additionally, since this method uses sample splitting and a traditional logistic mixed model, we can also use traditional M-estimation procedures (Stefanski and Boos 2002, Perez-Heydrich et al. 2014, Liu et al. 2016, Saul 2017) to estimate the asymptotic variance of these estimators on the testing sample. Inference is under-emphasized in related literature on the robustness to model misspecification assuming no interference, and this procedure provides a powerful tool to investigate it in the presence of partial interference.

2.6 Simulations

We investigated the finite-sample performance of the proposed method in a simulation study. We generated $D = 500$ datasets, where each dataset was comprised of $i = 1, \ldots, M = 800$ cluster. Each cluster had $j = 1, \ldots, N_i = 5$ individuals. The data generating process was very similar to those presented in Hajjem et al. (2017). For each individual, 8 pre-treatment covariates $L_{ij1}, \ldots, L_{ij8}$ were simulated from a correlated Gaussian distribution. The observed covariates were then transformed into six distinct regions - as in a decision tree - based on the values of $L_{ij1}, \ldots, L_{ij5}$ but irrespective of L_{ij6}, L_{ij7}, or L_{ij8}. That is, defining the partition $\Theta_0 = \{r_1, \ldots, r_T\}$ with $T = 6$, the transformed $\Theta_0(L_{ij}) = X_{ij}$ was a vector of $T = 6$ indicator variables, where $X_{ijt} = I(L_{ij} \in r_t)$. The form of the transformed covariates was not reflected in the observed data.

The probability that A_{ij} was equal to 1 depended logit-linearly on the individual's transformed covariates as well as a random intercept for the cluster:

$$\Pr(A_{ij} = 1 \,|\, L_{ij}, b) = \mathcal{L}^{-1}\Big(\Theta_0(L_{ij})^\intercal \theta_A + b\Big),$$

where $\mathcal{L}^{-1}(\theta_A) = (0.2, 0.4, 0.7, 0.3, 0.6, 0.8)^\intercal$. The observed values of the first 5 covariates thus had a non-logit-linear relationship with treatment, whereas the observed values of the last three covariates could be considered to be noise. Outcomes were generated under clustered interference, dependent on the individual's transformed covariates and treatment status as well as the average treatment status within the cluster:

$$\Pr(Y_{ij} = 1 \,|\, L_{ij}, A_i) = \mathcal{L}^{-1}\left(\Theta_0(L_{ij})^\intercal \theta_Y \,-\, 0.8A_{ij} \,-\, 0.6\sum_{j'=1}^{N_i} A_{ij'}/N_i\right),$$

where $\theta_Y = (0.8, 0.3, 0.4, 0.7, 0.2, 0.3)^\intercal$.

We employed GMERT-IPW using the stabilized estimators from Liu et al. (2016) for the inverse probability-weights. The GMERT algorithm was trained on $M_{train} = 150$ individuals, and the GMERT-LMM was fit to the remaining $M_{test} = 650$ clusters. We present two methods for comparison: The LMM-8 method fit a logistic mixed model that assumed that each of the 8 (untransformed) observed pre-treatment covariates had a logit-linear relationship with treatment. The LMM-5 method included only the correct 5 pre-treatment covariates, excluding the noise covariates, but still made a similar (incorrect) assumption that treatment related logit-linearly to the observed covariates. Since these latter two strategies relied on pre-specified models, they were fit to all $i = 1, \dots M$ clusters in the full sample $\mathcal{S} = \mathcal{S}_{train} \cup \mathcal{S}_{test}$.

Table 2.2: Estimates of $\mu(\alpha)$ for varying α from the three methods across the D datasets. Bias_{10^4} indicates the average bias times 10^4, and MSE_{10^4} indicates the average squared bias times 10^4; ASE_{10^3} is the average estimated (asymptotic) standard error times 10^3, ESE_{10^3} is the standard error of the estimates across the D datasets times 10^3, and SER equals ASE divided by ESE. Cov% is the empirical coverage of the Wald-type 95% confidence intervals. The proposed matching method performed well in all scenarios with low bias and nominal coverage rates. The traditional LMM-5 and LMM-8 methods experienced greater bias and less-than-nominal coverage.

Method	Target	Bias_{10^4}	MSE_{10^4}	Cov%	ASE_{10^3}	ESE_{10^3}	SER
GMERT	$\mu(0.45) = 0.464$	-7.72	2.50	95.0%	15.11	15.82	0.96
	$\mu(0.50) = 0.448$	-13.30	2.56	94.0%	15.09	15.95	0.95
	$\mu(0.55) = 0.431$	-15.37	2.57	93.6%	15.12	15.98	0.95
LMM-5	$\mu(0.45) = 0.464$	99.60	1.77	82.4%	9.31	8.82	1.06
	$\mu(0.50) = 0.448$	92.32	1.64	84.2%	9.18	8.89	1.03
	$\mu(0.55) = 0.431$	86.81	1.57	85.0%	9.13	9.02	1.01
LMM-8	$\mu(0.45) = 0.464$	66.80	1.37	90.0%	9.93	9.61	1.03
	$\mu(0.50) = 0.448$	58.76	1.28	91.4%	9.81	9.69	1.01
	$\mu(0.55) = 0.431$	52.95	1.24	90.6%	9.81	9.78	1.00

Table 2.2 presents summary statistics for estimates of $\mu(\alpha)$ from each of the three methods across all D datasets in the simulation study. The GMERT-IPW method performed well, with negligible bias and MSE. The SER was near 1, indicating that the average estimated standard errors from the sandwich variance estimator performed

well in relation to the standard error of the point estimates across the D datasets. The empirical coverage probability was almost exactly the nominal 95%.

Each of the LMM-5 and LMM-8 methods experienced a large amount of upward bias. Each method had a SER near 1, but failed to reach nominal coverage rates, likely due to the high bias. The LMM-8 method exhibited less bias and better coverage than LMM-5. The LMM-8 method may have been better in this scenario because including the latter 3 "noise" covariates may have allowed for more accurate estimates of cluster propensity scores, whereas the LMM-5 method was under-parameterized. The LMM-5 and LMM-8 methods, which relied on pre-specified yet incorrect modeling assumptions, were outperformed by GMERT-IPW in this scenario.

Table 2.3 presents summary statistics for the GMERT-IPW estimates of several Overall Effects. The LMM-5 and LMM-8 methods are not presented here since they were shown to perform poorly in Table 2.2. The proposed method experienced low bias and small MSE for these estimands. The sandwich variance estimator performed well, with SER near 1. The GMERT-IPW method again achieved nominal 95% coverage in this scenario.

Table 2.3: Summary of GMERT-IPW estimates of $\text{OE}(\alpha, \alpha')$ for varying allocations α'. The GMERT-IPW performed well for Overall Effects, with low bias and nominal coverage rates.

Target	Bias_{10^4}	MSE_{10^4}	Cov%	ASE_{10^3}	ESE_{10^3}	SER
$\text{OE}(0.55, 0.450) = -0.033$	-7.64	0.26	94.0%	5.14	5.01	1.03
$\text{OE}(0.55, 0.525) = -0.008$	-3.34	0.02	94.4%	1.30	1.28	1.02
$\text{OE}(0.55, 0.500) = -0.016$	-2.07	0.06	94.4%	2.59	2.54	1.02

2.7 Data Analysis

We illustrate the proposed methods in an analysis of a large-scale study of cholera vaccine effectiveness in Matlab, Bangladesh. At the beginning of the study, over 100,000

women and children from 6,415 baris (households of patrilineally-related individuals) were eligible to participate in a trial, in which each individual was randomized with equal probability to one of three treatment arms: placebo, B subunit-killed whole cell oral cholera vaccine, or killed whole cell-only oral cholera vaccine (Ali et al. 2005; 2009). Endpoint data was collected on all women and children, even those who did not choose to participate in the trial. After 1 year of follow up, anindividual was considered to have experienced the outcome if they had been infected by cholera during the year (i.e., coded as $Y = 1$), and was considered to have not experienced the outcome if they had not experienced cholera infection during the year (Y=0). Since endpoint data exists for all individuals, and since participation in the experimental component of the study was not controlled by the study's design, potential for confounding exists when analyzing the data.

As in Perez-Heydrich et al. (2014) we consider an individual to have been treated if they had received at least two doses of either of the active cholera vaccines ($A = 1$), and to have been untreated otherwise ($A = 0$). Partial interference was assumed at the level of the bari as in Barkley et al. (2017), as there is evidence that transmission of cholera often takes place within-bari (Ali et al. 2005).

It is assumed that conditioning on the pre-treatment covariates for age (centered, in decades) and distance to nearest river (in kilometers) is sufficient to satisfy conditional exchangeability, as well as positivity. To estimate cluster propensity scores, the dependent variable was trial participation, where $B = 1$ indicated that an individual presented for inclusion in the randomized component of the study, and $B = 0$ otherwise. Estimates of the cluster propensity scores for the procedures can easily be obtained by including a randomization probability of 2/3 as in Perez-Heydrich et al. (2014). We employ the the unstabilized estimators in Section 2.4 because the stabilized estimators from Liu et al. (2016) target different estimands since cluster sizes vary.

We briefly describe the cross-validation procedure used to select tuning parameters; further details for setup and analysis of the Matlab cholera vaccine study are presented in Section A.1. To select the tuning parameters of the decision tree within the GMERT algorithm, we carry out $K = 5$-fold cross-validation on a training sample of the data. Small clusters with $N_i < 5$ were included in the testing sample, and the largest cluster with $N_i = 244$ individuals was included in the training sample; otherwise, all clusters were assigned at random to the training sample with probability $1/4$ or the testing sample with probability $3/4$. After partitioning into the two samples, there were $M_{train} = 1,465$ clusters representing $31,395$ individuals in $\mathcal{S}_{train}$, and $M_{test} = 4,950$ clusters representing $90,580$ individuals in $\mathcal{S}_{test}$. Summary statistics for bari size, and treatment and participation within bari stratified by the two samples are provided in Section A.1.2. The training sample seems to be representative of the testing sample. This indicates that the GMERT algorithm learned on the training sample may be able to help provide reasonable inferences on the testing sample of data.

To carry out cross-validation, for each of the K folds, the training sample was partitioned into distinct sub-samples. That is, for $k = 1, \ldots, K$, partition $\mathcal{S}_{train} = \mathcal{S}_{tr,k} \dot{\cup} \mathcal{S}_{cv,k}$, where $\mathcal{S}_A \dot{\cup} \mathcal{S}_B$ indicates that $\mathcal{S}_A \cap \mathcal{S}_B = \emptyset$. This partition was carried out so that each cluster i had at least one individual in $\mathcal{S}_{tr,k}$ and at least one other individual in $\mathcal{S}_{cv,k}$. This was done so that empirical BLUPs from each cluster could be estimated for each individual in $\mathcal{S}_{cv,k}$, as our interest is in appropriately capturing the correlation between individuals within clusters, which is analogous to repeated measures for a single subject in traditional longitudinal data settings (Chen et al. 2015). Further details are provided in A.1.2. Summary statistics indicate that these cross-validation folds should be satisfactory for selecting appropriate tuning parameters.

The following values were considered for tuning parameters: maximum tree depth

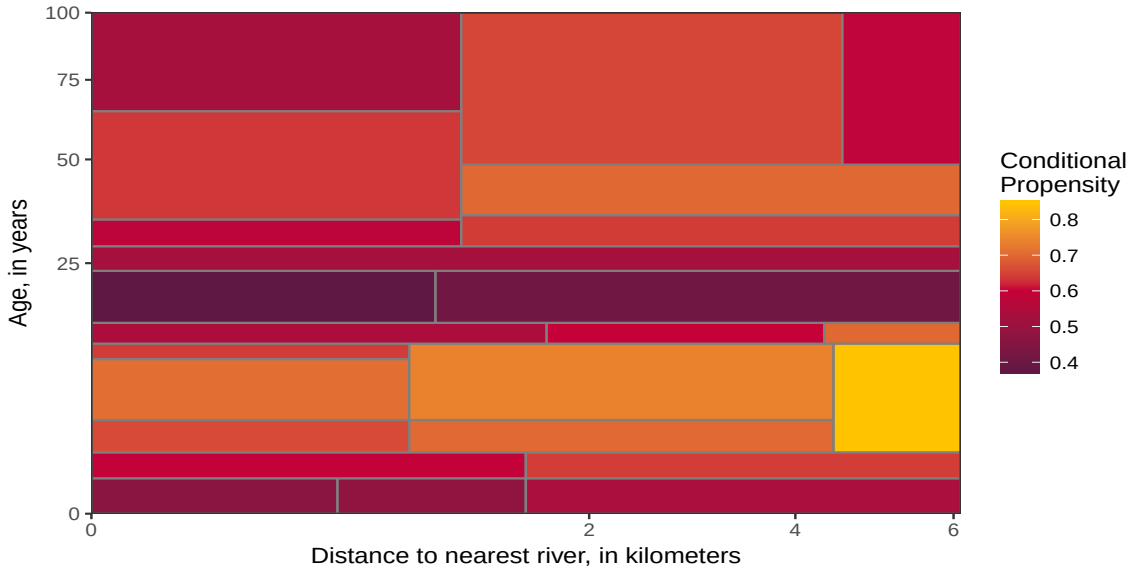
equal to 5 or 15, minimum observations to split a node equal to 50 or 25, and minimum observations in a terminal node equal to 10 or 30 (see Table A.1). For each of the proposed 6 sets of tuning parameters, we applied the GMERT algorithm to $\mathcal{S}_{tr,k}$ and then used the resulting parameter estimates (fixed effects from the tree, as well as empirical BLUPs) to predict the response (participation) of the individuals in $\mathcal{S}_{cv,k}$. This was repeated with all models and all $k = 1, \ldots, K$. Weighted misclassification error was chosen for the loss function to determine the best model. For each model, summary statistics of the loss over the K folds were calculated, and the tuning parameters corresponding to the GMERT model with lowest mean loss were chosen to be the final tuning parameters (see Table A.2). Further details are provided in Section A.1.3.

The selected tuning parameters were: maximum tree depth equal to 5, minimum observations to split a node equal to 50, and minimum observations in a terminal node equal to 30. The first two tuning parameters were the same as suggested by Hajjem et al. (2017). The third tuning parameter was larger than the suggested 10, indicating that the GMERT algorithm had better prediction in this scenario when terminal nodes were larger. The GMERT algorithm was then applied on the entire training set, $\mathcal{S}_{train}$, using these selected tuning parameters.

The resulting tree was recovered from this final run of the GMERT algorithm for use in propensity score estimation on the testing set. The tree had 24 terminal nodes, partitioning the space of L_{age} and L_{dist} into 24 regions. As described in Section 2.5, this tree was used as a model selection procedure on the testing set. That is, the covariates L_{age} and L_{dist} were transformed into a series of indicator variables X_i corresponding to the 24 regions defined by terminal tree nodes. Then, a logistic mixed effects model was fit for participation with a random intercept for bari, and fixed effects for each region (i.e., 24 categorical fixed effects, one random intercept, and no fixed effect intercept).

A comparison of the estimated fixed effects for this tree in the two samples is provided in Figure A.4. This comparison of the predicted probabilities from the training sample and the estimated probabilities resulting from the logistic mixed model from the testing sample suggests at most mild overfitting of the data, as the two sets of probabilities are quite similar for nearly all of the 24 tree nodes. Requiring a minimum of 30 observations in each terminal node of the tree likely resulted in having larger terminal nodes. That is, perhaps fewer terminal nodes could be characterized as containing a small group of spurious extreme values, which would mitigate overfitting of the algorithm (Athey and Imbens 2016).

Figure 2.1: A representation of the estimated fixed effects from the logistic mixed model used in the GMERT-IPW method. The covariate space is shown partitioned as according to the tree, and colored according to the estimated value of the corresponding fixed effect from the GMERT-LMM estimated on the training dataset.

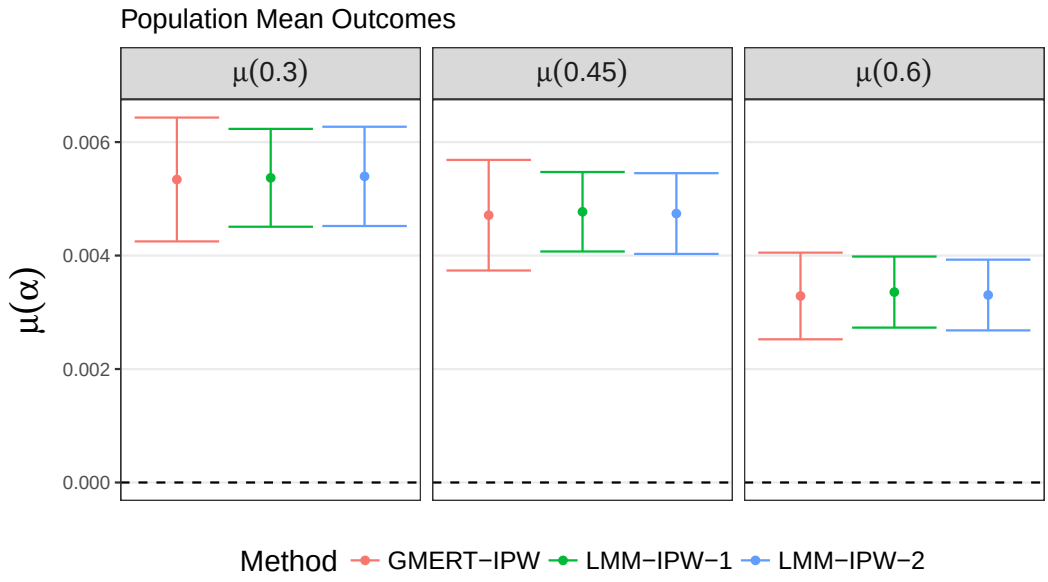


A visual representation of the fitted logistic mixed model in the GMERT-IPW method is presented in Figure 2.1. The 24 regions of the predictor space are indicated in this figure, and colored according to the estimated fixed effect parameter from the logistic mixed model; the variance component for the random intercept was estimated to be $\hat{\sigma}^2 = 2.82$. These model estimates are used for propensity score estimation in

the GMERT-IPW procedure, and will likely result in cluster propensity score estimates that differ from those in the comparator methods that use more traditional modeling techniques.

For comparison to the GMERT-IPW procedure, we also fit two versions of the IPW estimators using more traditional logistic mixed effects modeling assumptions. The first method, labeled LMM-IPW-1, includes linear terms for individual age and distance to river; the second method, LMM-IPW-2, includes linear terms and an interaction for age and distance to river, as well as a quadratic term for age. Each of these models is fit with a single random intercept for bari. These modeling assumptions were pre-specified and thus were fit to the entire data sample, $\mathcal{S} = \mathcal{S}_{train} \cup \mathcal{S}_{test}$.

Figure 2.2: Estimates and 95% CIs for population mean estimands $\mu(\alpha)$ from the cholera vaccine study.

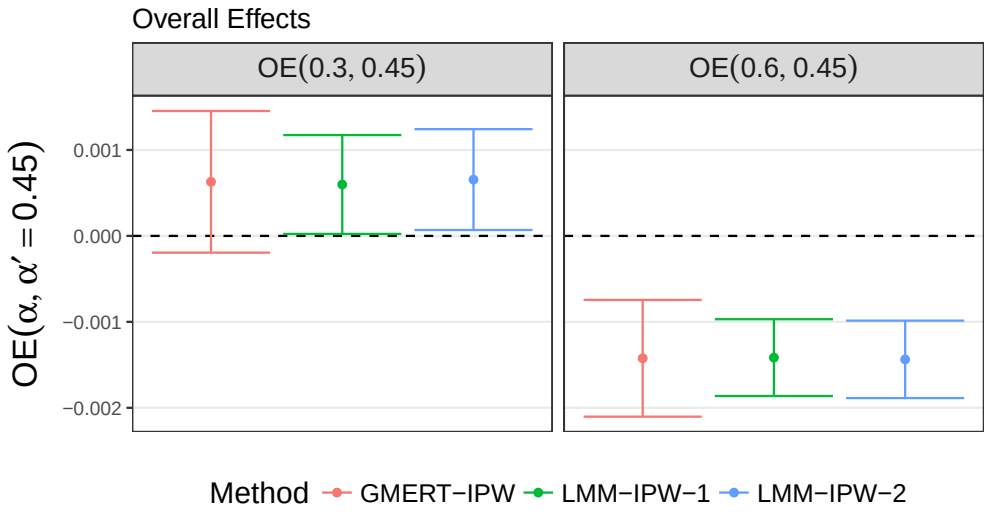


We now compare the estimates of the causal parameters from the GMERT-IPW estimator to those from two LMM-IPW methods that used more traditional, pre-specified modeling assumptions. Figure 2.2 presents the estimates of $\mu(\alpha)$ from the different models. Estimates from GMERT-IPW indicated a rate of cholera infection of about

5.5 cases per 1000 person-years when $\alpha = 0.3$. This rate was lower when the treatment allocation α increased to 0.45 and 0.6. It appears there could be a non-linear decrease in the rate of cholera infection as α increases.

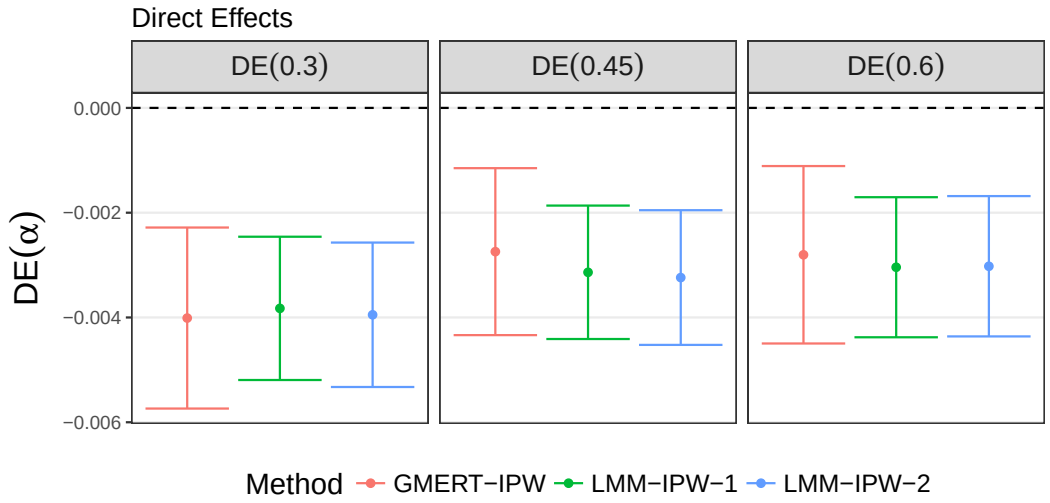
The point estimates were very similar for all three methods. The 95% CIs were wider for the GMERT-IPW method than for either of the two methods using traditional modeling techniques. One contributing factor is that GMERT-IPW is fit on only M_{test} clusters, or about 75% of the sample size of the LMM-IPW methods that are fit to all M clusters. It's also likely that the GMERT-IPW 95% CIs are wider because the GMERT-LMM model requires estimation of 25 parameters, whereas the model for LMM-IPW-1 estimates 4 parameters, and 6 for LMM-IPW-2. Estimates of the

Figure 2.3: Estimates and 95% CIs for Overall Effects from the cholera vaccine study.



Overall Effects are presented in Figure 2.3. The point estimates of all three method are positive for $\hat{\text{OE}}(0.3, 0.45)$ and negative for $\hat{\text{OE}}(0.6, 0.45)$, indicating a favorable effect of higher levels of treatment allocation. The GMERT-IPW method estimates $\hat{\text{OE}}(0.6, 0.45) = -0.0015$, indicating a reduction in the number of cases of cholera by 1.5 per 1000 person-years when increasing the treatment allocation from $\alpha' = 0.45$ to $\alpha = 0.6$. The corresponding 95% CI excludes zero by a notable margin, indicating a

Figure 2.4: Estimates and 95% CIs for Direct Effects from the cholera vaccine study.



significant and protective effect of this increase in treatment allocation. In contrast, the GMERT-IPW method estimates $\hat{\text{OE}}(0.3, 0.45)$ to be less than 1 and the corresponding 95% CI includes zero by a non-negligible margin, which fails to indicate a significant effect. This provides further evidence of a non-linear relationship between cholera outcomes and treatment allocation strategy, and perhaps suggests some amount of herd protectiveness.

The point estimates of each of the Overall Effects presented here were similar across the three methods. However, the 95% CIs were again wider for the GMERT-IPW method than for the LMM-IPW methods. In fact, the 95% CIs for each of the LMM-IPW methods excluded zero for $\text{OE}(0.3, 0.45)$, albeit by a small margin, so these methods estimated a statistically significant effect here where the GMERT-IPW method did not. This is an example of where the proposed GMERT-IPW method results in different inference about causal effects than existing methods.

Point estimates for Direct Effects are presented in Figure 2.4. Each method indicates a significant and protective Direct Effect of treatment at each level of α. For the GMERT-IPW method, the estimate of $\hat{\text{DE}}(0.6)$ is smaller in magnitude than the

estimate of $\hat{\text{DE}}(0.3)$; again, this suggests a non-linear relationship between treatment allocation strategy and cholera outcomes. Point estimates are similar across the GMERT-IPW and LMM-IPW methods for each of the three Direct Effects shown here. Further results, including those for the Indirect and Total Effects, are provided in Section A.1.4.

2.8 Discussion

Performance of existing estimators for the estimands in Tchetgen Tchetgen and VanderWeele (2012) relies on the correct specification of a logistic mixed model. This paper introduces a new method that may reduce bias due to model mis-specification in this setting, where partial interference and correlated treatment are plausible. The proposed method involves an application of the GMERT algorithm (Hajjem et al. 2017), combining machine learning and traditional mixed modeling techniques, to reduce the amount of pre-specified modeling assumptions. The proposed method exhibits much less bias than existing methods in simulations. Wald-type confidence intervals constructed from the sandwich variance estimator reach nominal 95% coverage in the simulation study, again outperforming existing techniques.

The proposed methods are illustrated in a study of cholera vaccination in over 100,000 individuals. The estimators showed some evidence that increased vaccine allocation reduces the risk of of infection by cholera. The proposed methods indicated a non-linear relationship between vaccine allocation, α, and risk of infection, $\mu(\alpha)$. For example, the Direct Effect of vaccination was estimated to be greater at $\alpha = 0.3$ than at $\alpha = 0.6$. The proposed methods found relatively strong evidence of favorable Overall and Indirect Effects of increasing treatment allocation from $\alpha' = 0.45$ to $\alpha = 0.6$, but did not find strong evidence of favorable Overall or Indirect effects of increasing treatment allocation from $\alpha' = 0.3$ to $\alpha = 0.45$. Taken together, this may suggest some amount of herd protectiveness. In contrast, existing estimators found strong evidence

of favorable Overall Effects in both cases. These results may help to inform decisions on vaccination strategies, especially when treatment is scarce or expensive.

The proposed method uses sample-splitting to ensure regularity conditions for valid statistical inference, which is relatively under-emphasized in some related research on robust modeling methods for propensity score estimation. Future research should consider adjusting the node-splitting rules to account for the honest estimation procedure; e.g., see Athey and Imbens (2016). Different machine learning methods should be investigated for future extensions of the GMERT-IPW method. For example, the MARS technique (Friedman 1991, Milborrow 2014) may provide improvements over decision trees. Alternatives to GMERT should also be investigated, e.g., the (mixed) model-based partitioning algorithm introduced in Fokkema et al. (2017). Relatively parsimonious methods like these can take advantage of existing software and methods for inference on the test sample.

In the absence of interference, machine learning methods for estimating propensity scores generally focus on covariate balance between the treated and untreated groups; see e.g., Pirracchio and Carone (2016) or Austin and Stuart (2015). Whether this notion generalizes to estimating cluster propensity scores assuming partial interference is an open question. Replacing decision trees with ensemble learners would likely exhibit better predictive accuracy for GMERT (for example, compare Hajjem et al. (2011) and Hajjem et al. (2014) for analogous work with continuous responses), which may result in an extension of GMERT-IPW that is even more robust to model mis-specification. Robust modeling techniques like these may provide some of the tools towards estimating optimal treatment regimes from observational studies in the presence of interference (Zhao et al. 2012, Zhang et al. 2012).

This paper assumes a treatment model with a single random intercept with respect to the cluster identifier that follows a Normal distribution. The assumed form of the

cluster propensity score in (2.19) could be generalized to integrate over the distribution of multiple random effects. Notably, the GMERT algorithm can be carried out with more than one random effect. Generalizing (2.19) would extend these methods to more complex treatment model correlation structures, and should be considered for future research.

CHAPTER 3: A MATCHING ESTIMATOR FOR CLUSTERED INTERFERENCE

3.1 Introduction

Inferring causal effects from an observational study is challenging because participants are not randomized to treatment. Traditional associative statistical methods cannot adequately control for confounding, and are subject to bias. Specialized methods exist for drawing inferences about causal parameters from non-experimental data; for example, inverse probability-weighting (IPW) and matching methods are commonly used for this purpose.

Infectious disease research often exhibits the additional challenge of interference, which is when one individual's outcome may be affected by another individual's treatment. Popular causal parameters in this setting are found in Tchetgen Tchetgen and VanderWeele (2012), and several IPW methods have been proposed for drawing inference to those causal parameters from non-randomized studies exhibiting some interference (see e.g., Perez-Heydrich et al. (2014) and Liu et al. (2016)). A drawback of IPW estimators is that in general they can experience instability when weights grow large, which we show is exacerbated in this setting.

We introduce a covariate matching method to estimate causal parameters in Tchetgen Tchetgen and VanderWeele (2012). These methods extend related work from Abadie and Imbens (2006) and Yang et al. (2016) to a setting where some types of interference is possible. The remainder of the paper is as follows. A brief overview of matching and imputation in causal inference is presented in Section 3.2. In Section 3.3

we describe the causal parameters of interest and the assumed structure of interference. The estimators are proposed in Section 3.4. In Section 3.5 the estimators are shown to be consistent and asymptotically Normal, and we present an estimator for the asymptotic variance. Extensions are considered in Section 3.6, and existing IPW estimators are discussed in Section 3.7. Finite-sample properties of the estimators are investigated with a simulation study in Section 3.8. The proposed estimators are then illustrated in Section 3.9 with an analysis of a regional cholera vaccine study from Matlab, Bangladesh. The main paper concludes with a discussion.

3.2 Matching Methods for Causal Inference

In the potential outcomes framework, causal effects of treatment are defined by differences in potential outcomes that may follow treatment events. The causal effect of treatment for a study unit is the difference in the health outcome that the unit would have had if it had experienced treatment, and the health outcome that the unit would have had if it had not experienced treatment. For example, consider a data sample of $i = 1, \ldots, N$ study units where $A_i = 1$ indicates that the study unit was observed to experience treatment and $A_i = 0$ indicates that they were not, and where $Y_i = 1$ indicates that unit i was observed to experience an unfavorable outcome following treatment and $Y_i = 0$ indicates that they were not. In one possible scenario, the i^{th} study unit is observed to experience treatment $A_i = 1$ and then experiences an antecedent health status $Y_i(a = 1)$. In another possible scenario, the i^{th} study unit instead does not undergo treatment and then experiences an antecedent health status $Y_i(a = 0)$. The quantities $Y_i(a)$ for $a = 0, 1$ are called potential outcomes, and the causal effect of treatment for the i^{th} unit is defined to be $Y_i(1) - Y_i(0)$. Since only $Y_i(a = A_i)$ is observed and $Y_i(a = 1 - A_i)$ is not, the individual causal effect is inherently unobservable; however it is estimable from observable data under certain assumptions.

The fundamental assumption for estimating causal parameters from observational studies is that observed treatment is independent of potential outcomes after conditioning on a set of potential confounders (Rubin 1978). Using the notation of Dawid (1979), this assumption states that there is a sufficient set of covariates L_i such that conditional exchangeability holds:

$$Y_i(a) \perp A_i \mid L_i \text{ for all units } i \text{ and for } a = 0, 1. \tag{3.23}$$

Covariate matching methods obtain estimates of causal effects by pairing two study units when they have (nearly) identical covariates L but different observed treatments (i.e., units i and i' with $L_i \approx L_{i'}$ but $A_i \neq A_{i'}$) and then perhaps taking the difference of their observed outcomes (e.g., $Y_i - Y_{i'}$). Given their similarity to the causal effects and fundamental assumption presented above, matching methods are widely considered to be easily interpretable, and are very popular; an excellent review is presented in Stuart (2010). Neither IPW nor matching methods is superior in all scenarios; see e.g., Frölich (2004a), Busso et al. (2014) and Huber et al. (2013). Unlike IPW, relatively few matching methods have been introduced for scenarios where multiple treatment levels exist; some examples include Lechner (2001), Foster (2003), Frölich (2004b), Feng et al. (2012), McCaffrey et al. (2013), Rassen et al. (2013), Fong and Imai (2014) and Yang et al. (2016). Rarer still are matching methods in the presence of interference.

Individual causal effects can also be estimated by imputing all unobserved potential outcomes. Although imputation often relies on explicit modeling assumptions, this paper uses a covariate matching procedure to impute potential outcomes, as in Abadie and Imbens (2006). We define $j^* = \arg\min_{\{j\,:\,A_j = 1 - A_i\}} ||X_i - X_j||$ to be the "matching" unit with most-similar covariates to unit i, and then impute $\tilde{Y}_i(1 - A_i) = Y_{j^*}$. In the case of binary treatment, then the individual causal effect could be estimated by the quantity $Y_i - \tilde{Y}_i(1 - A_i)$, times $(-1)^{1-A_i}$.

The proposed methods follow such a "matched-imputation" procedure. Yang et al. (2016) extended matched-imputation to the setting where multiple treatment levels exist. Here, matched-imputation is extended to the setting where interference may be possible within distinct clusters of individuals. We consider each cluster to be a study unit. We propose methods for matching clusters to one another to impute cluster-level potential outcomes in order to estimate causal parameters in Tchetgen Tchetgen and VanderWeele (2012). Asymptotic properties, as well as an estimator for the asymptotic variance, draw from Abadie and Imbens (2006).

3.3 Study Setup and Existing Estimands

Consider a super-population of distinct clusters of individuals; for ease of notation, index the clusters by i. In general, clusters can have varying numbers of individuals; we consider the case where each cluster has the same number of individuals and $C_i = c$ for some $c \geq 2$. Let $A_i = (A_{i1}, \ldots, A_{ic})$ be the ordered vector of treatment statuses for all individuals in the cluster i, where $A_{ij} = 1$ indicates that individual j in the cluster obtained treatment. Similarly, let $Y_i = (Y_{i1}, \ldots, Y_{ic})$ be the ordered vector of observed outcomes for the cluster, where Y_{ij} might indicate whether or not individual j in the cluster was observed to become infected. We assume that the ordering of individuals within groups is non-informative.

Let $\mathcal{A}(c)$ be the set of all binary vectors of length c such that $a \in \mathcal{A}(c)$ is a potential treatment vector for any cluster of c individuals. It is assumed that the outcome of an individual may be affected by the treatment of any other individual in the same cluster, but not by treatments of individuals from different clusters. This has been termed partial interference (Sobel 2006) or clustered interference (Barkley et al. 2017), and it is assumed here and in the sequel. With the assumption of clustered interference, we can define potential outcomes with respect to the treatments in $\mathcal{A}(c)$. For $a \in \mathcal{A}(c)$,

define $Y_{ij}(a)$ to be the potential outcome for the j^{th} individual in cluster i when the cluster experiences treatment a. Define $Y_i(a) = (Y_{i1}(a), Y_{i2}(a), \ldots, Y_{ic}(a))$ to be the potential outcome for cluster i with respect to treatment a for any $a \in \mathcal{A}(c)$. By counterfactual consistency, the cluster's observed outcome is $Y_i = Y_i(A_i)$. Define the average cluster-level potential outcome with respect to a to be $\overline{Y}_i(a) = c^{-1} \sum_{j=1}^{c} Y_{ij}(a)$, such that the average observed outcome is $\overline{Y}_i = c^{-1} \sum_{j=1}^{c} Y_{ij}$.

The estimands of interest presented here are generalizations of those in Tchetgen Tchetgen and VanderWeele (2012), which defines causal estimands in the presence of clustered interference that arise from the setting of a two-stage cluster-randomized trial, similar to Hudgens and Halloran (2008). In such a trial, each cluster is assigned to a treatment arm that corresponds to a parameter, e.g., α or α', where the parameter governs the distribution of treatment within the cluster. Tchetgen Tchetgen and VanderWeele (2012) considers trials that follow a "type B parameterisation": when a cluster is assigned to the treatment arm corresponding to $\alpha \in [0, 1]$, then each individual in that cluster is exposed to treatment independently with probability α. A causal effect of interest is the contrast in expected potential outcomes when clusters are assigned to the α arm or the α' arm.

For a single cluster of c individuals, there are $|\mathcal{A}(c)| = 2^c$ potential unique treatment vectors for the cluster. Under the type B parameterization, the probability that the cluster would experience treatment $a \in \mathcal{A}(c)$ equals the Bernoulli-type product $\pi(a, \alpha) = \prod_{j=1}^{c} \alpha^{a_j}(1-\alpha)^{1-a_j}$. Thus, the expected average potential outcome for a cluster under this type B parameterization is $\sum_{a \in \mathcal{A}(c)} \bar{Y}_i(a)\pi(a, \alpha)$. An estimand of interest is the (population) mean potential outcome in the α arm:

$$\theta(\alpha) = \mathbb{E}\left[\sum_{a \in \mathcal{A}(c)} \bar{Y}_i(a)\pi(a, \alpha)\right]. \tag{3.24}$$

Here, $\mathbb{E}(\cdot)$ indicates that the average is taken over all clusters in the super-population. The Overall Effects are differences in two population mean potential outcomes:

$$\theta(\alpha, \alpha') = \theta(\alpha) - \theta(\alpha'). \tag{3.25}$$

3.4 Proposed Matched-Imputation Estimators

We extend the matched-imputation estimator from works including Imbens (2000), Abadie and Imbens (2006) and Yang et al. (2016) to the setting of clustered interference in order to estimate (3.24) and (3.25). Let there be a sample of $i = 1, \ldots, N$ clusters, and let each cluster contain c individuals. Let each cluster have an i.i.d. copy of the observed variables (L_i, A_i, Y_i). As above, $Y_i = (Y_{i1}, \ldots, Y_{ic})$ is the ordered vector of observed outcomes, where $Y_{ij} = 1$ might indicate that individual j became infected by the end of follow-up. Observed values of treatment are indicated by the ordered vector $A_i = (A_{i1}, \ldots, A_{ic})$, where $A_{ij} = 1$ indicates that individual j in the cluster was administered treatment, e.g., vaccinated.

Baseline (or pre-treatment) covariates for all individuals in the cluster are included in L_i. In the case where one covariate is measured for each individual, $L_i = (L_{i1}, \ldots, L_{ic})$ is a $c \times 1$ column vector, where L_{ij} indicates the (scalar) value of the single covariate for individual j in the cluster. In the case where there are $p > 1$ covariates measured for each individual, then $L_i = (L_{i1}, \ldots, L_{ic})$ is then a $c \times p$ matrix, where $L_{ij} = (L_{ij1}, \ldots, L_{ijp})$ is a $1 \times p$ row vector of the p ordered covariates for individual j in the i^{th} cluster. We denote by $L_{i \cdot p} = (L_{i1p}, \ldots, L_{icp})$ the $c \times 1$ column vector that contains information on the p^{th} covariate for all individuals in the cluster.

For positivity, we assume that

$$\Pr(A_i = a \mid L_i) > 0 \text{ for all } L_i \text{ and any } a \in \mathcal{A}(c). \tag{3.26}$$

We assume that the ordering of individuals within each cluster is not informative. We also assume that there is at least one cluster with observed treatment $A_i = a$ for all $a \in \mathcal{A}(c)$. We assume a version of weak conditional exchangeability with respect to the cluster average potential outcome:

$$\bar{Y}_i(a) \perp I(A_i = a) \mid L_i \text{ for all clusters } i \text{ and for all } a \in \mathcal{A}(c). \tag{3.27}$$

As described in Yang et al. (2016), assuming (3.27) and (3.26) allow for estimation of the population mean outcomes for each treatment. Matched-imputation can be employed to estimate the following quantities for each $a \in \mathcal{A}(c)$:

$$\mathbb{E}[\bar{Y}_i(a)] = \mathbb{E}\left[\mathbb{E}\left(\bar{Y}_i | A_i = a, L_i\right)\right].$$

Multiplying the estimate of the above quantity by $\pi(a, \alpha)$, and then summing over all $|\mathcal{A}(c)| = 2^c$ of these products allows for estimation of $\theta(\alpha)$ and $\theta(\alpha, \alpha')$. Next, we introduce a simple method for transforming each unit's pre-treatment covariates L_i into a row-vector in order to easily carry out the matching procedure.

We propose re-arranging the covariate structure such that the $c \times p$ matrix L_i is transformed into a row vector X_i that preserves all information about the cluster's covariates. In the base case when $p = 1$, define $X_i = L_i$. When $p > 1$, then define $X_i = (L_{i\cdot 1}, L_{i\cdot 2}, \ldots, L_{i\cdot p})$. That is, X_i is a $(p * c) \times 1$ row vector, where the first c terms are the vector of observed values of the 1st covariate (ordered by the order of individuals in the cluster), and the second c terms are the vector of observed values of the 2nd covariate, and so on. After these transformations, we assume analogues of positivity (3.26) and weak conditional exchangeability (3.27) hold on X_i. We consider matching based on Mahalanobis distance $||X - X'||$. Note that we can define $T = |\mathcal{A}(c)|$ and re-label the unique treatment vectors $a \in \mathcal{A}(c)$ from w_1 to w_T to arrive at a

notation consistent with the case of matching with multiple treatments in the case of no interference as in Yang et al. (2016).

We proceed by taking each unit i and, for each of the 2^c-1 unobserved levels of treatment for that unit, find a "matching" unit with that level of treatment whose covariates are close. That is, let unit i have observed values $(X_i, A_i, \bar{Y}_i)$, and let $a \in \mathcal{A}(c)$ such that $a \neq A_i$, and let $\mathcal{S}_a = \{j | A_j = a\}$ be the set of all units whose observed treatment equals a. Define $j_1(i,a) = \arg\min_{j \in \mathcal{S}_a} ||X_i - X_j||$ to be the unit in whose covariates are the most similar to those of unit i (among all units with observed treatment $A = a$). Extending this notation, the 2^{nd} closest match to unit i in $\mathcal{S}_a$ is indicated by $j_2(i,a)$, and the m^{th} closest match is indicated by $j_m(i,a)$. Denote $J_M(i,a) = \{j_1(i,a), \ldots, j_M(i,a)\}$ to be the set of the $M \geq 1$ nearest matches to unit i with respect to treatment $a \neq A_i$.

For some value of M pre-specified by the investigator, the imputed cluster average potential outcome is the mean of the outcomes for the units in $J_M(i,a)$:

$$\tilde{Y}_i(a) = M^{-1} \sum_{m=1}^{M} \bar{Y}_{j_m(i,a)}. \tag{3.28}$$

In a special case, define $\tilde{Y}_i(A_i) = \bar{Y}_i$. Define $K_M(i) = \sum_{l=1}^{N} I(i \in J_M(l, A_i))$ to be the number of times that unit i is matched to other units after completing the matching procedure on a sample; this quantity plays a role in variance estimation. The $N * M * (2^c - 1)$ imputed potential outcomes are used to estimate $\theta(\alpha)$ and $\theta(\alpha, \alpha')$, respectively, with the following formulas:

$$\hat{\theta}_M(\alpha) = \frac{1}{N} \sum_{i=1}^{N} \left[\sum_{a \in \mathcal{A}(c)} \tilde{Y}_i(a) \pi(a, \alpha) \right] \tag{3.29}$$

$$\hat{\theta}_M(\alpha, \alpha') = \frac{1}{N} \sum_{i=1}^{N} \left[\sum_{a \in \mathcal{A}(c)} \tilde{Y}_i(a) \big\{ \pi(a, \alpha) - \pi(a, \alpha') \big\} \right]. \tag{3.30}$$

3.5 Asymptotic Properties

We follow Abadie and Imbens (2006) in showing that, under certain assumptions, the proposed estimators are consistent and asymptotically Normal. We also propose an estimator for the asymptotic variance following Abadie and Imbens (2006) instead of relying on the bootstrap, which may not be appropriate (Abadie and Imbens 2008). Many details are left until later sections; note that the pre-treatment covariates are usually expressed after being transformed from L_i to the row-vector of X_i.

Define $\mu(x,a) = \mathbb{E}(\bar{Y}|A=a, X=x)$ and $\mu_a(x) = \mathbb{E}(\bar{Y}(a)|X=x)$. The variance of the average outcomes, conditional on $X=x$, is $\sigma^2(x,a) = \text{Var}(\bar{Y}|A=a, X=x)$, and its causal counterpart is $\sigma_a^2(x) = \text{Var}(\bar{Y}(a)|X=x)$. Residuals with respect to $\mu_A(X)$ are $\epsilon_i = \bar{Y}_i - \mu_{A_i}(X_i) = \bar{Y}_i - \mathbb{E}(\bar{Y}_i(A_i)|X=X_i)$.

We decompose the estimator for the population mean estimands into three components (see Section B.1.1):

$$\hat{\theta}_M(\alpha) = \overline{\theta(\alpha|X)} + E_M(\alpha) + B_M(\alpha). \tag{3.31}$$

This decomposition is key for the asymptotic properties shown in this section; the three terms on the right side are as follows. The sample average conditional mean outcome $\overline{\theta(\alpha|X)}$ relates to the expected potential outcomes over the sample of N units,

$$\overline{\theta(\alpha|X)} = \frac{1}{N}\sum_{i=1}^{N}\sum_{a\in\mathcal{A}(c)}\mu_a(X_i)\pi(a,\alpha). \tag{3.32}$$

The second term, $E_M(\alpha)$ is a weighted average of residuals,

$$E_M(\alpha) = \frac{1}{N}\sum_{i=1}^{N}\left(1+\frac{K_M(i)}{M}\right)\epsilon_i. \tag{3.33}$$

Finally, $B_M(\alpha)$ is a conditional bias term (relative to $\overline{\theta(\alpha|X)}$):

$$B_M(\alpha) = \frac{1}{N}\frac{1}{M}\sum_{i=1}^{N}\sum_{m=1}^{M}\sum_{a\in\mathcal{A}(c)}\left[\mu_a(X_{j_m(i,a)}) - \mu_a(X_i)\right]\pi(a,\alpha)I(A_i \neq a). \tag{3.34}$$

Asymptotic consistency of the estimators is established in Theorem B.1.2: we show that $\overline{\theta(\alpha|X)} \xrightarrow{p} \theta(\alpha)$, and that $E_M(\alpha) = o_p(1)$. Let k be the number of continuous pre-treatment covariates (with respect to the transformed X). When $k = 1$ then $N^{1/2}(\hat{\theta}_M(\alpha) - \theta(\alpha))$ is $o_p(1)$ and the estimator is asymptotically Normal. When $k > 1$ then the conditional bias may not vanish and the estimator may not be asymptotically Normal. We present a proof of asymptotic Normality that ignores the conditional bias term, similar to Theorem 4 in Abadie and Imbens (2006).

Theorem 3.5.1. ***Asymptotic normality of the matching estimator:*** *Assume, as in Abadie and Imbens (2006, Assumption 1), that X is a random vector having density that is bounded and bounded away from zero on* $\text{supp}(X) = \mathbb{X} \subset \mathbb{R}^k$. *Assume a random sample of i.i.d. clusters of individuals. Assume weak conditional exchangeability and positivity on the (transformed) pre-treatment covariates, X. Furthermore, assume the following smoothness conditions from Abadie and Imbens (2006, Assumption 4) for all $a \in \mathcal{A}(c)$: $\mu(x,a)$ and $\sigma^2(x,a)$ are Lipschitz in $\mathbb{X}$, $\mathbb{E}(\bar{Y}^4|A = a, X = x)$ exists and is bounded uniformly in x, and $\sigma^2(x,a)$ us bounded away from zero. Then*

$$\left(V^{E_M(\alpha)} + V^{\theta(\alpha|X)}\right)^{-1/2} N^{1/2}\left(\hat{\theta}_M(\alpha) - B_M(\alpha) - \theta(\alpha)\right) \xrightarrow{d} N(0,1).$$

Proof. Write the sample average conditional mean outcome as $\overline{\theta(\alpha|X)} = \frac{1}{N}\sum_{i=1}^{N}\theta(\alpha|X_i)$. Here, $\mathbb{E}[\theta(\alpha|X_i)] = \theta(\alpha)$. Defining

$$V^{\theta(\alpha|X)} = \text{Var}\left[\theta(\alpha|X_i)\right] = \text{Var}\left[\sum_{a\in\mathcal{A}(c)}\mu_a(X_i)\pi(a,\alpha)\right], \tag{3.35}$$

then it is straightforward that

$$\left(V^{\theta(\alpha|X)}\right)^{-1/2} N^{1/2} \left(\overline{\theta(\alpha|X)} - \theta(\alpha)\right) \xrightarrow{d} N(0,1). \tag{3.36}$$

Now, define the variance component relating to $E_M(\alpha)$ to be

$$V^{E_M(\alpha)} = \frac{1}{N} \sum_{i=1}^{N} \left(1 + \frac{K_M(i)}{M}\right)^2 \pi(A_i, \alpha)^2 \sigma^2(X_i, A_i). \tag{3.37}$$

In Theorem B.1.3 we use the Lindeberg-Feller Central Limit Theorem on the conditional distribution of $E_M(\alpha)$ to show that

$$\left(V^{E_M(\alpha)}\right)^{-1/2} N^{1/2} E_M(\alpha) \xrightarrow{d} N(0,1). \tag{3.38}$$

Asymptotic independence of the terms in (3.36) and (3.38) completes the proof. □

Define $V_M^{\alpha} = V^{E_M(\alpha)} + V^{\theta(\alpha|X)}$ to be the variance term for $\hat{\theta}_M(\alpha) - B_M(\alpha) - \theta(\alpha)$ found in Theorem 3.5.1. An estimator for $V^{E_M(\alpha)}$ is

$$\hat{V}_J^{E_M(\alpha)} = \frac{1}{N} \sum_{i=1}^{N} \left(1 + \frac{K_M(i)}{M}\right)^2 \pi(A_i, \alpha)^2 \hat{\sigma}_J^2(X_i, A_i),$$

which relies on the within-treatment-level matching estimator $\hat{\sigma}_J^2(X_i, A_i)$ introduced in Abadie and Imbens (2006) (see Section B.2.1). We estimate $V^{\theta(\alpha|X)}$ with:

$$\hat{V}_{M,J}^{\theta(\alpha|X)} = \frac{1}{N} \sum_{i=1}^{N} \left(\sum_{a \in \mathcal{A}(c)} \tilde{Y}_i(a) \pi(a, \alpha) - \hat{\theta}_M(\alpha) \right)^2 + \frac{1}{N} \sum_{i=1}^{N} \left(1 + \frac{K_M(i)}{M^2}\right) \pi(A_i, \alpha)^2 \hat{\sigma}_J^2(X_i, A_i).$$

Relevant details are provided in Sections B.2.2 and B.2.3.

We estimate $\hat{V}^{\alpha}_{M,J}$ by summing $\hat{V}^{E_M(\alpha)}_{J}$ and $\hat{V}^{\theta(\alpha|X)}_{M,J}$ to arrive at:

$$\hat{V}^{\alpha}_{M,J} = \frac{1}{N}\sum_{i=1}^{N}\left(\sum_{a\in\mathcal{A}(c)}\tilde{Y}_i(a)\pi(a,\alpha) - \hat{\theta}_M(\alpha)\right)^2 + \frac{1}{N}\sum_{i=1}^{N}K^{**}_M(i)\pi(A_i,\alpha)^2\hat{\sigma}^2_J(X_i,A_i).$$

Here, $K^{**}_M(i) = (K_M(i)/M)^2 + (2M - 1/M)(K_M(i)/M)$ is the multiplicative factor arising from the number $K_M(i)$ of times unit i is used as a match. Theorem 7 in Abadie and Imbens (2006) suggests that $\hat{V}^{\alpha}_{M,J}$ may be a consistent estimator for $\mathrm{Var}(\hat{\theta}_M(\alpha))$ under certain scenarios. With respect to the Overall Effects, we similarly propose to estimate the asymptotic variance of $\hat{\theta}_M(\alpha,\alpha')$ with:

$$\hat{V}^{\alpha,\alpha'}_{M,J} = \frac{1}{N}\sum_{i=1}^{N}\left(\sum_{a\in\mathcal{A}(c)}\tilde{Y}_i(a)\big\{\pi(a,\alpha) - \pi(a,\alpha')\big\} - \hat{\theta}_M(\alpha,\alpha')\right)^2 + \frac{1}{N}\sum_{i=1}^{N}K^{**}_M(i)\big\{\pi(a,\alpha) - \pi(a,\alpha')\big\}^2\hat{\sigma}^2_J(X_i,A_i) \tag{3.39}$$

3.6 Extensions

Reducing the dimension of treatment from 2^c may provide practical improvements under certain scenarios. We discuss some potential extensions here.

Define $f_\Sigma(a) = \sum_j^c a_j$ to be the function summing all elements of a vector. Assuming that the ordering of individuals within clusters is uninformative implies that $\mathbb{E}\left[\bar{Y}_i(a)\right] = \mathbb{E}\left[\bar{Y}_i(a')\right]$ whenever $f_\Sigma(a) = f_\Sigma(a')$. We propose a version of the matched-imputation estimator that assumes treatment irrelevance within strata formed by the sum of the within-cluster treatment. That is, we assume there are only $c+1$ unique treatment values (up to ordering of the treatment vector a) defined by $f_\Sigma(a) = 0$, $\ldots$, $f_\Sigma(a) = c$. This reduces the number of unique treatments from 2^c to c, and thus the number of matches necessary. The estimators are carried out by using $\pi_\Sigma(s,\alpha) = \sum_{a\in\mathcal{A}(s)}\pi(a,\alpha)I(f_\Sigma(a) = s)$, with the corresponding changes in the variance estimators.

We investigate these estimators in the simulation study in Section 3.8, and implement them in the data analysis in Section 3.9.

The estimators that assume treatment irrelevance within strata may perform well by lowering variance at low cost in terms of bias. Further gains may be possible by considering stronger assumptions. An investigator could assume there is a sequence of cutoff values $0 < \xi_1 < \cdots < \xi_{T-1} < 1$ such that $\mathbb{E}\left[\bar{Y}_i(a)\right] = \mathbb{E}\left[\bar{Y}_i(a')\right]$ whenever $\xi_t < f_\Sigma(a), f_\Sigma(a') \leq \xi_{t'}$. This method may provide noticeable benefit when c is large, especially when compared to IPW methods that are often unstable in such a setting.

We now present two potential extensions for the case when cluster sizes may vary. First, consider the scenario when the data are not sparse and there exists at least one cluster with observed treatment equal to a for all $a \in \bigcup_i \{\mathcal{A}(C_i)\}$. In this simpler case, it would be possible to carry out any of the above estimators within strata defined by cluster size C_i; that is, matching only clusters of the same size to each other. All components necessary for estimation (and variance estimation) are obtained.

This method may also be extended to the case where clusters of different sizes are matched to each other. There may be settings in which two clusters may contain unequal numbers of individuals, but the two clusters may also be reasonably considered to be similar enough to match to one another. In this setting, one could consider an assumption of treatment irrelevance within strata defined by cutoffs $\{\xi_t\}$. The information-preserving transformation of covariates from L_i to X_i described above may not be appropriate when clusters have different sizes; in this setting or simply when clusters have a large amount of individuals, it may be appropriate to assume that there is a set of within-cluster summary statistics of L_i that is sufficient for satisfying conditional exchangeability.

3.7 Comparator: Existing Inverse Probability Weighted Estimators

Tchetgen Tchetgen and VanderWeele (2012) propose inverse probability of treatment weighted estimators for inference which are in the form of the Horvitz-Thompson estimators (Horvitz and Thompson 1952, Rosenbaum 1987). For (3.24), these are:

$$\hat{\theta}_{HT}(\alpha) = N^{-1} \sum_{i=1}^{N} \frac{\bar{Y}_i \pi(A_i, \alpha)}{\Pr(A_i = a | L_i = l_i)}$$

where the conditional probability in the denominator is the cluster propensity score (CPS). The conditional independence assumption is

$$Y_i(a) \perp A_i \mid L_i \text{ for any cluster } i \text{ and any vector } a \in \mathcal{A}(c),$$

and the CPS is estimated following a logistic mixed model for treatment A on covariates L with a random intercept for cluster identifier. Positivity (i.e., (3.26)) is also assumed. The estimator for the target effects is

$$\hat{\theta}_{HT}(\alpha, \alpha') = \hat{\theta}_{HT}(\alpha) - \hat{\theta}_{HT}(\alpha') = N^{-1} \sum_{i=1}^{N} \frac{\bar{Y}_i [\pi(A_i, \alpha) - \pi(A_i, \alpha')]}{\Pr(A_i = a | L_i = l_i)}$$

Estimates of the asymptotic variance of these estimators is traditionally obtained via M-estimation theory and the sandwich variance estimator; see e.g. Perez-Heydrich et al. (2014) or Saul and Hudgens (2017). We call these the "IPW-HT" estimators.

Liu et al. (2016) proposed a set stabilized IPW estimators for similar estimands, following the Hajek form. In the special case when all clusters have the same number of units, the estimands that these estimators target are identical to the estimands of interest here. In contrast to the unstabilized IPW-HT estimators above, we refer to

these stabilized estimators the "IPW-Hajek" estimators:

$$\hat{\theta}_{Hajek}(\alpha) = \hat{N}^{-1} \sum_{i=1}^{N} \frac{\bar{Y}_i \pi(A_i, \alpha)}{\Pr(A_i = a | L_i = l_i)}$$

$$\hat{\theta}_{Hajek}(\alpha, \alpha') = \hat{\theta}_{Hajek}(\alpha) - \hat{\theta}_{Hajek}(\alpha')$$

where $\hat{N}$ is now replaced with an estimated term. Liu et al. (2016) proposes two methods for this term, and we employ the second method:

$$\hat{N} = \sum_{i=1}^{N} \frac{\pi(A_i, \alpha)}{\Pr(A_i = a | L_i = l_i)}.$$

The IPW-Hajek estimators follow similar assumptions (i.e., (3.7) and (3.26)) and parameteric models, and can use M-estimation theory (Stefanski and Boos 2002, Saul and Hudgens 2017) to obtain estimates of the asymptotic variance via the sandwich technique for the scenarios we are interested in.

3.8 Simulation Study

Four simulation studies were carried out to determine the finite-sample performance of the proposed matching estimators. Each simulation study was carried out for a fixed cluster size, i.e., $c = 3, 5, 8$, and 14. Otherwise, each of the four simulation studies was carried out in the same fashion. For each simulation study, we generated $D = 300$ datasets where each dataset contains $i = 1, \ldots, N = 500$ clusters. Each cluster was fixed to have $C_i = c$ individuals, where c varied by simulation study. Each cluster had an i.i.d. copy of the observed variables (L_i, A_i, Y_i), where A_i was a binary vector indicating observed treatment for the c individuals and Y_i was the binary vector indicating observed outcome for the c individuals. Here L_i was a $(c \times 3)$-dimensional matrix of pre-treatment variables, where $L_{ij} = (L_{ij1}, L_{ij2}, L_{ij3})$ was the row vector of

the observed values for individual j, and $L_{i\cdot p} = (L_{i1p}, L_{i2p}, L_{i3p})^\intercal$ was the column vector of the observed values of the p^{th} covariate for each of the individuals in the cluster.

For any individual j in any cluster i, L_{ij1} and L_{ij2} were Bernoulli variables and equal to 1 with probability 0.75. The third covariate L_{ij3} was a categorical variable that took on the levels v_1, v_2 and v_3 with equal probability. The probability that A_{ij} was equal to 1 depended logit-linearly on these observed covariates as well as a random intercept for the cluster:

$$\begin{aligned}\Pr(A_{ij} = 1 | L_{ij}, b) = \mathcal{L}^{-1}(\, -2.4 + 0.2I(L_{ij3} = v_2) \, + \, 0.2I(L_{ij3} = v_3) \, + \\ 1.25L_{ij1} \, + \, 0.75L_{ij2} \, + \, 0.95L_{ij1}L_{ij2} \, + \, b \,)\end{aligned}$$

Outcomes were generated under clustered interference, dependent on the individual's pre-treatment covariates and treatment status as well as the average treatment status within the cluster:

$$\begin{aligned}\Pr(Y_{ij} = 1 | A_i, L_{ij}) = \mathcal{L}^{-1}(\, -0.1 - 0.15I(L_{ij3} = v_2) \, - \, 0.35I(L_{ij3} = v_3) \, + \\ 0.15L_{ij1} \, + \, 0.2L_{ij2} \, - \, 1.25A_{ij} \, - \, 1.05 \sum_{j'=1}^{c} A_{ij'}/c \,)\end{aligned}$$

The true estimands were determined empirically using 10,000 clusters with covariates generated as above. Then, for each treatment $a \in \mathcal{A}$, the potential outcome $Y_{ij}(a)$ for each individual in every cluster was generated using the causal analogue $\Pr(Y_{ij}(a) = 1 | L_{ij})$ to $\Pr(Y_{ij} = 1 | A_i, L_{ij})$ from above. These potential outcomes were then combined with the corresponding values of $\pi(a, \alpha)$, and sums and means were taken to obtain the true values of $\theta(\alpha)$ as in (3.24) and $\theta(\alpha, \alpha')$ as in (3.25).

We fit the proposed matching estimator to each dataset, reducing the dimension of treatment by assuming treatment irrelevance within strata formed by the sum of

treated individuals as described in Section 3.6. To create a dataset with one row per i.i.d. study unit, we transformed each cluster's $(c \times 3)$-dimensional matrix L_i of pre-treatment covariates into a $(1 \times 3 * c)$-dimensional row vector of covariates X_i as in Section 3.4. Conditional exchangeability and positivity hold conditional on X_i for the $c+1$ treatment levels.

We specified $M = 1$ so that each cluster was matched to c other units (one for each level of treatment except for the observed level of treatment). The variance estimators were carried out by specifying $J = 1$ within-treatment matches to estimate $\sigma^2(A_i, X_i)$, as described in Section B.2.1. That is, results are shown for the estimators $\hat{\theta}_{M=1}(\alpha)$ and $\hat{\theta}_{M=1}(\alpha, \alpha')$, where their asymptotic variances are estimated with $\hat{V}^{\alpha}_{M=1,J=1}$ and $\hat{V}^{\alpha,\alpha'}_{M=1,J=1}$, respectively. Wald-type 95% confidence intervals are presented. Matching was performed on the covariates, without modeling.

For comparison, we also fit the IPW-Hajek estimators introduced in Liu et al. (2016). as described in In this case, where each cluster has c individuals, these estimators target estimate the same estimands. We fit the IPW-Hajek estimators with the correctly specified logistic mixed model for treatment with a random intercept for cluster, main effects for each of the 3 pre-treatment covariates, and an interaction between the two binary pre-treatment covariates. Wald-type 95% CIs for these estimators are calculated from the sandwich variance estimators. These stabilized estimators performed much better in finite samples than the IPW-HT from Tchetgen Tchetgen and VanderWeele (2012), which are not shown here.

In Table 3.4, we present summary statistics of the performance of the estimators targeting $\theta(\alpha)$ across the $D = 300$ datasets. The proposed matching estimator performs very well. For the point estimates, there is negligble bias and very small MSE. The proposed variance estimator also seems to perform well, with SER near 1 indicating that the estimated standard errors are about appropriate for the method. The empirical

Table 3.4: Estimates of $\theta(\alpha)$ for varying α and the two methods across the four simulation scenarios. Each of the four scenarios is run with a fixed cluster size, c. Bias_{10^4} indicates the average bias times 10^4, and MSE_{10^4} indicates the average squared bias times 10^4; ASE_{10^3} is the average estimated (asymptotic) standard error times 10^3, ESE_{10^3} is the standard error of the estimates across the D datasets times 10^3, and SER equals ASE divided by ESE. Cov% is the empirical coverage of the Wald-type 95% confidence intervals. The proposed matching method performs well in all scenarios with low bias and nominal coverage rates. The IPW-Hajek method performs well for $c = 3$ but experiences increasing MSE and decreasing coverage as cluster sizes increase.

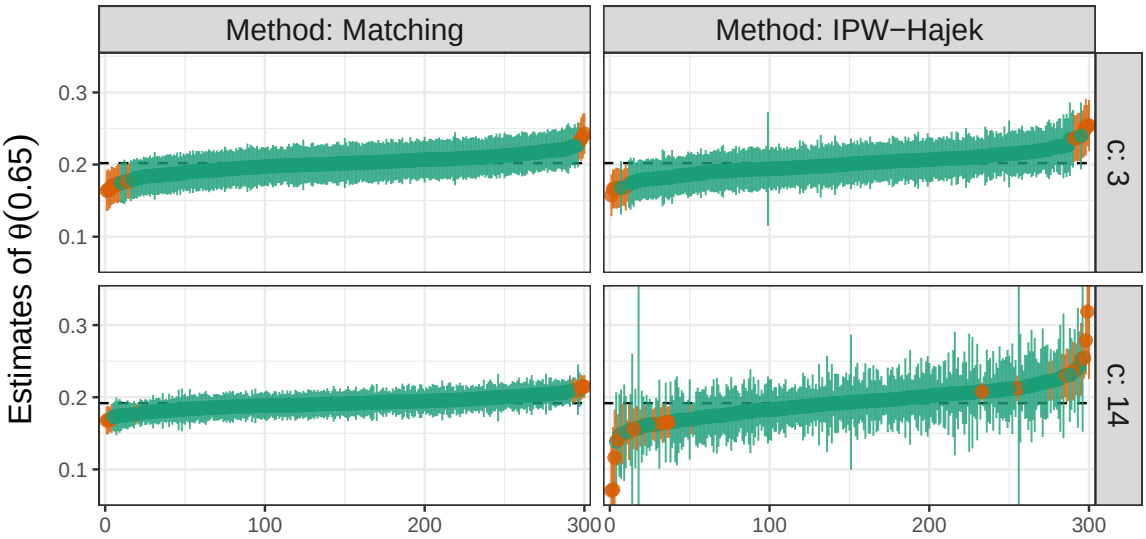
c	Target	Method	Bias_{10^4}	MSE_{10^4}	Cov%	ASE_{10^3}	ESE_{10^3}	SER
3	$\theta(0.50)$=0.258	Match	4.38	1.87	95.0%	14.02	13.69	1.02
		IPW	10.75	3.05	93.3%	16.88	17.46	0.97
	$\theta(0.65)$=0.202	Match	-8.98	1.66	95.0%	14.00	12.86	1.09
		IPW	-9.87	2.56	93.7%	15.53	16.00	0.97
5	$\theta(0.50)$=0.252	Match	18.49	1.49	94.7%	11.76	12.08	0.97
		IPW	8.82	3.89	92.7%	18.24	19.74	0.92
	$\theta(0.65)$=0.197	Match	0.31	1.21	96.7%	11.75	11.04	1.06
		IPW	-15.20	3.09	92.7%	16.83	17.54	0.96
8	$\theta(0.50)$=0.250	Match	13.90	1.15	94.7%	10.56	10.64	0.99
		IPW	8.48	6.01	92.0%	19.95	24.53	0.81
	$\theta(0.65)$=0.193	Match	12.91	1.10	94.7%	10.70	10.41	1.03
		IPW	2.64	4.81	90.0%	18.11	21.98	0.82
14	$\theta(0.50)$=0.248	Match	-3.48	0.92	94.3%	9.60	9.58	1.00
		IPW	-11.11	7.09	89.7%	23.84	26.65	0.89
	$\theta(0.65)$=0.192	Match	-2.57	0.78	96.3%	9.71	8.84	1.10
		IPW	6.09	7.27	89.0%	21.63	27.00	0.80

coverage probability of the Wald-type 95% CIs is approximately the nominal 95%.

The IPW-Hajek estimator performs well when cluster sizes are smallest. For example, when $c = 3$ then the IPW-Hajek estimator experiences fairly low bias, SER close to 1, and empirical coverage percentages near the nominal 95%. However, as cluster sizes increase, then the IPW-Hajek estimator experiences strinctly increasing MSE, strictly increasing ASE and ESE, and strictly worsening empirical coverage.

We also present a "forest plot" in Figure 3.5 to illustrate the performance of these methods. The vertical lines represent the 95% Wald-type confidence intervals corresponding to each point estimate. The simulations are re-ordered so that the point

Figure 3.5: Forest plot of estimates of $\theta(0.65)$ for the two methods across two simulation studies. The dotted black horizontal line indicates the true value of $\theta(0.65)$ in each simulation study. Illustrated in each panel are $D = 300$ point estimates and corresponding 95% CIs; one for each simulated dataset. The reddish color indicates that the 95% CI excludes the true value of the parameter, and the greenish color indicates that the 95% CI includes the true value of the parameter. The proposed methods (left side) perform well in both scenarios. IPW-Hajek method (right side) exhibits instability and behaves erratically when cluster size increases.



estimates are increasing from left-to-right for each panel. The horizontal, dotted black line indicates the true value of $\theta(\alpha = 0.65)$, and each estimate and corresponding confidence interval has a greenish color when the confidence interval includes the true value (and reddish when it does not). In the left panels, the proposed matching estimator is illustrated to perform quite well in both scenarios shown, for $c = 3$ and $c = 14$. In the right panels, the IPW-Hajek estimator performs well for $c = 3$, but exhibits erratic behavior when $c = 14$.

In Table 3.5, we present summary statistics of the performance of the proposed estimators of the Overall Effects, $\theta(\alpha, \alpha')$. The MSE is very low. The average bias is negligible: at worst, in the third row, the average bias is 31.4% of the average standard error, and coverage is still good here at 93%. In all cases, the SER is near 1 indicating

that the asymptotic variance estimator is performing well for this estimator. The empirical coverage percent ranges from 93% to 97.3%, quite close to the nominal 95%.

Table 3.5: Estimates of $\theta(\alpha, \alpha')$ from the proposed matching methods in each of the four simulation scenarios. Bias_{10^4} indicates the average bias times 10^4, and MSE_{10^4} indicates the average squared bias times 10^4; ASE_{10^3} is the average estimated (asymptotic) standard error times 10^3, ESE_{10^3} is the standard error of the estimates across the D datasets times 10^3, and SER equals ASE divided by ESE. Cov% is the empirical coverage of the Wald-type 95% confidence intervals. The proposed matching method performs well in all scenarios with low bias and nominal coverage rates.

c	Target	Bias_{10^4}	MSE_{10^4}	Cov%	ASE_{10^3}	ESE_{10^3}	SER
3	$\theta(0.65, 0.5) = -0.056$	-13.36	0.28	96.3%	5.58	5.16	1.08
	$\theta(0.7, 0.45) = -0.095$	-5.63	0.72	97.3%	9.18	8.47	1.08
5	$\theta(0.65, 0.5) = -0.055$	-18.17	0.37	93.0%	5.78	5.79	1.00
	$\theta(0.7, 0.45) = -0.095$	3.51	0.87	93.7%	9.31	9.35	1.00
8	$\theta(0.65, 0.5) = -0.057$	-1.00	0.40	93.3%	6.44	6.30	1.02
	$\theta(0.7, 0.45) = -0.093$	-13.65	0.98	93.0%	10.08	9.84	1.02
14	$\theta(0.65, 0.5) = -0.057$	0.91	0.50	94.3%	7.30	7.10	1.03
	$\theta(0.7, 0.45) = -0.094$	2.67	1.11	94.0%	10.82	10.55	1.03

The IPW-Hajek estimators for Overall Effects are not shown in Table 3.5 as their behavior for this target mirrors their behavior illustrated in Table 3.4. In particular, as cluster sizes increase, the IPW-Hajek estimators suffers from increasing MSE and decreasing empirical coverage rates to below nominal levels. The proposed methods thus perform well in multiple scenarios for both $\theta(\alpha)$ and $\theta(\alpha, \alpha')$ with respect to multiple allocations α, α', and outperform existing IPW estimators.

3.9 Data Analysis

The proposed matching estimator is illustrated with an analysis of a cholera vaccine study in Matlab, Bangladesh (Ali et al. 2005; 2009). Over 100,000 women and children from 6,415 baris (i.e., households of patrilineally-related individuals) were eligible to participate in the study. There was an experimental and a non-experimental component

to the study, and each eligible individual could choose to participate. Participants were randomized with equal probability to one of three treatment arms: B subunit-killed whole cell oral cholera vaccine, killed whole cell-only oral cholera vaccine, or placebo. Non-participants did not receive either version of active treatment. Since individuals could choose to participate in the study, there is a non-neglible non-experimental component to the study, and potential for confounding exists when analyzing the endpoint data. An individual was indicated to have experienced the outcome if they had been infected with cholera by the end of 1 year; these individuals were coded as having outcome 1, and the uninfected individuals were coded as having outcome 0, so lesser values of the target estimands represent more favorable health outcomes.

We consider any individual to have received treatment if they were vaccinated with at least two doses of one of the two cholera vaccines as in Perez-Heydrich et al. (2014); otherwise the individual was considered to be untreated. We assume partial or clustered interference at the level of the bari as in Barkley et al. (2017), as evidence exists that transmission of cholera often takes place within baris (Ali et al. 2005). To illustrate this method, we subset the data to clusters of size 3. That is, the 211 baris containing only 1 or 2 individuals were excluded from the analysis, and exactly 3 individuals were chosen to represent baris with more than 3 individuals where the remaining individuals from those baris were also excluded. Figure 3.6 presents the empirical distribution of treatment for the N =6,194 baris of size 3 (for a total of 18,582 individuals) that remain after the subsetting process.

We illustrate the matching method assuming treatment irrelevance for the sum of individuals treated as described in Section 3.6 and as investigated in the simulation study in Section 3.8. For these estimators, positivity and weak conditional exchangeability were assumed to hold conditional on individual age L_{ij1} (centered, in decades) and distance to river L_{ij2} (in kilometers). Since all individuals within the same bari

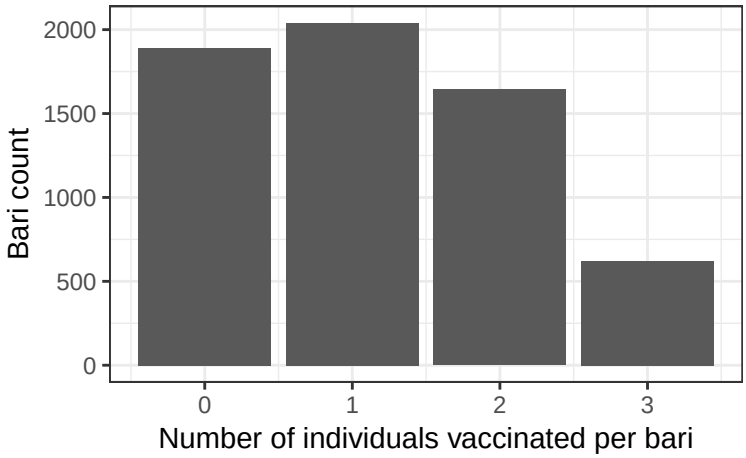


Figure 3.6: Empirical distribution of the number of individuals vaccinated per cluster after subsetting the data

were equidistant to river, the data transformation produced a 1×4 matrix X_i where $X_{ij} = L_{ij1}$ for $j = 1, 2, 3$ and X_{i4} equals the distance from bari to river. The matching procedure was carried out on this 1×4 vector X_i of pre-treatment covariates for all $i \leq N$ clusters. The estimators were fit with $M = 1$ pre-specified, thereby requiring $c = 3$ matching units per cluster. Wald-type 95% CI's were constructed from the proposed asymptotic variance estimator, which was carried out using $J = 1$ within-treatment-level matches to estimate the conditional variance term. We refer to this method Match-1.

We fit additional versions of the proposed matching estimator by specifying varying values of M and J. When matching discrepancy is non-negligible, using a greater number of matches in either the point estimation procedure or variance estimation procedure may result in better estimates from either procedure. Shown below are results from the estimator where we specified $M = 3$ for the estimator and $J = 3$ for the variance estimator, labeled Match-2.

The IPW-Hajek estimators from Liu et al. (2016) were fit for comparison. Positivity and conditional exchangeability were assumed conditional on age (centered) and

distance to river. A logistic mixed effects model was fit with a random intercept for each cluster, a linear term for distance to river, and linear and quadratic terms for age. Wald-type 95% CI's were constructed from the empirical sandwich variance estimator. A second set of IPW-Hajek estimators was fit, where the model this time included linear and quadratic terms for age and distance to river, as well as an interaction between their linear terms. These estimators are labeled IPW-Hajek-1 and IPW-Hajek-2, respectively. IPW-HT estimates are not shown, as they performed very similarly to the IPW-Hajek estimators.

Figure 3.7: Estimates and confidence intervals of $\theta(\alpha)$ for the Matlab cholera vaccine study. Each estimator is color-coded. The two versions of the proposed matching estimators are shown with circles, and the IPW-Hajek estimators are shown with triangles.

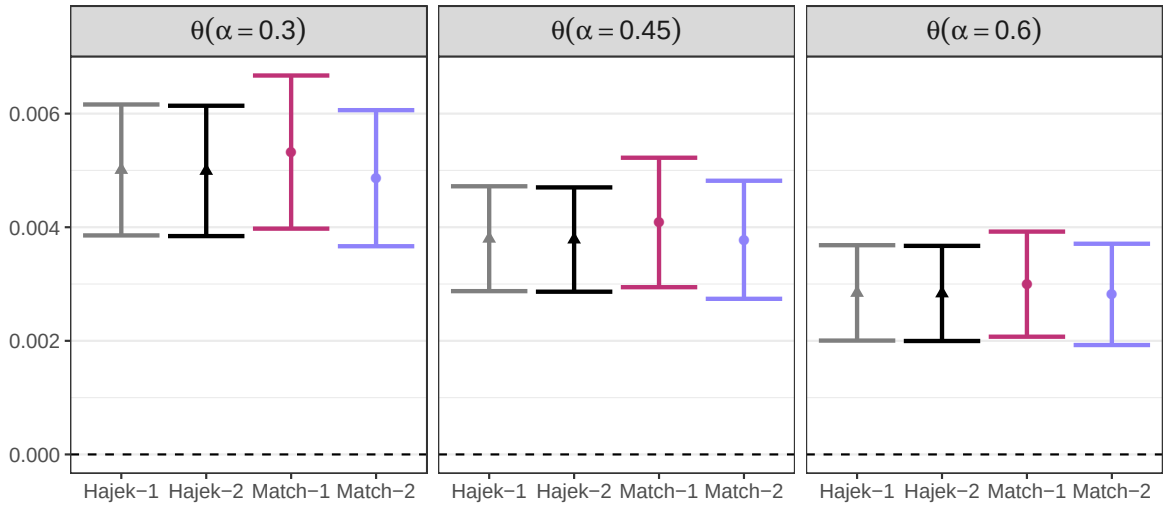
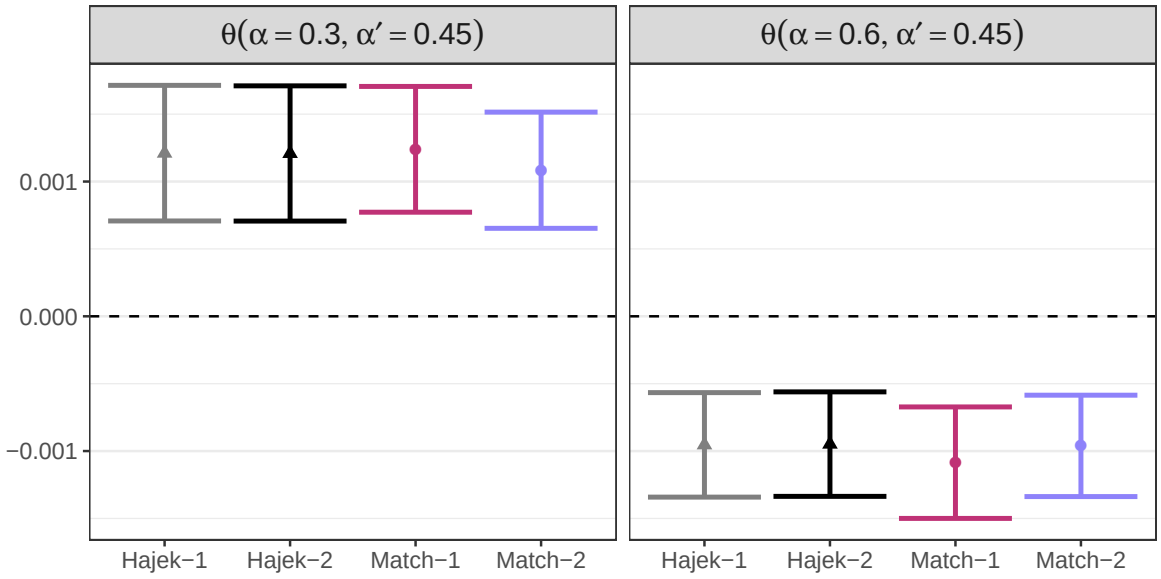


Figure 3.7 depicts point estimates and corresponding 95% CIs of the population mean estimands, $\theta(\alpha)$, for a range of levels of α from all four methods. The Match-1 method estimates decreasing values of $\mu(\alpha)$ as α increases across the three levels, suggesting that increasing treatment allocation α could result in a reduction in the rate of infection by cholera. Point estimates for the Match-2 method are slightly smaller in magnitude than those for Match-1; this may indicate that bias due to matching

Figure 3.8: Estimates and confidence intervals of Overall Effects for the Matlab cholera vaccine study. Each estimator is color-coded and labeled on the x-axis. In the right panel, these results indicate a reduction of the cholera infection rate by 1 case per 1000 individuals annually when increasing from α'=0.45 to $\alpha = 0.6$.



discrepancy is minimal. The point estimates for the two IPW-Hajek methods are very similar to one another, and also to those from the Match-2 estimator.

Figure 3.8 depicts point estimates and corresponding 95% CIs for the Overall Effect estimands, $\theta(\alpha, \alpha')$, for a range of levels of α and α'. Recall that lower values of $\theta(\alpha)$ are more favorable, and since $\theta(\alpha, \alpha')$ is defined to be the effect of switching from α' to α, negative estimates of the Overall Effect correspond to favorable health outcomes for the type B policy α compared to α'. The Match-1 method estimates an unfavorable effect for $\theta(0.3, 0.45)$ and a favorable effect for $\theta(0.6, 0.45)$. Each of these two estimates are statistically significant, as the 95% CI excludes zero in each case. For example, increasing the type B policy from $\alpha' = 0.45$ to $\alpha = 0.6$ is estimated to decrease the rate of cholera infection by about 1 case per 1000 individuals annually. The point estimate of the Match-2 estimator is slightly smaller in magnitude, and the width of the 95% CIs are also slightly smaller, but it performs very similarly to Match-1. The IPW-Hajek

estimators again performed similarly to the proposed methods, although the width of the 95% CIs for $\theta(0.3, 0.45)$ from the two IPW-Hajek methods are perhaps slightly greater than those from the matching methods.

3.10 Discussion

Introduced in this paper is a method for estimating causal effects in the presence of partial interference that relies on covariate matching. The estimators are consistent and asymptotically Normal in some scenarios, and estimators for the asymptotic variance are proposed. The proposed methods perform well in a simulation study, exhibiting low bias and reaching nominal coverage. These methods are illustrated on data collected from a large vaccine study. These methods estimated significant and favorable effects of greater treatment allocation on risk of infection by cholera. The results provide further strength of evidence that increasing vaccine allocations reduces the risk of infection by cholera, especially since the proposed methods do not rely on modeling assumptions.

Existing IPW methods for the target estimands rely on the correct specification of the treatment model and cluster propensity scores. Although these IPW estimators are at risk of bias due to model mis-specification, development of methods more robust to modeling assumptions remains an open problem (Liu and Hudgens 2014). The proposed matching estimators present a model-free method for drawing inference to these target estimands. The proposed methods outperform the existing IPW estimators in a finite-sample study, where the instability of the IPW estimators is shown to be exacerbated when cluster sizes increase. Matching methods are generally considered to be highly interpretable; the proposed estimators should be a valuable addition to the public health researcher's toolkit.

The proposed estimators rely on the untestable assumptions of conditional exchangeability and clustered interference. Future work should consider extending these

estimators to target the Indirect, Total, and Direct Effects in Tchetgen Tchetgen and VanderWeele (2012) or related estimands in Hudgens and Halloran (2008). Asymptotic Normality of the proposed estimators is not guaranteed when there are multiple continuous covariates per cluster. Another drawback of this method is that the number of covariates per study unit increases as the cluster size increases. As discussed in Section 3.6, future research may consider covariate dimension reduction strategies, which may also allow for matching clusters when the clusters do not contain the same number of individuals. Future research may also investigate the benefits of using modeling methods for propensity score matching as in Abadie and Imbens (2016).

There are few matching estimators that can estimate causal effects in the presence of interference from an observational study, yet such methods may provide valuable information (Arpino and Mattei 2016, Arpino et al. 2017). By defining one cluster of individuals to be a study unit, this paper extends the matched-imputation estimators in Abadie and Imbens (2006) and Yang et al. (2016) to the scenario of clustered interference. The estimators introduced here are the first that do not rely on IPW to target the estimands in Tchetgen Tchetgen and VanderWeele (2012) from an observational study. Furthermore, the analysis of the Matlab cholera vaccine study represents the first application of a matching method to estimate causal effects in the presence of interference from a non-experimental public health study.

CHAPTER 4: CAUSAL INFERENCE FROM OBSERVATIONAL STUDIES WITH CLUSTERED INTERFERENCE

4.1 Introduction

Inferring causal effects from an observational study (also referred to as a non-randomized or non-experimental study) is challenging because participants may select their own treatment. Observational studies in many settings such as infectious disease research present the additional challenge that one individual's treatment may have an effect on another individual's outcome, i.e., there may be interference (Cox 1958). For example, whether one individual is administered a vaccine may affect whether another individual develops disease from some infectious pathogen. In certain settings it may be reasonable to assume that individuals can be partitioned into clusters such that there may be interference among individuals within a single cluster, yet no interference between individuals in distinct clusters. Sobel (2006) described this assumption as "partial interference"; here this assumption is referred to as "clustered interference." Clusters might entail households, classrooms, geographical areas, or other hierarchical structures. Several types of treatment effects (i.e., causal estimands) have been proposed for the setting where there may be clustered interference; e.g., see Halloran and Struchiner (1995), Hudgens and Halloran (2008) and Tchetgen Tchetgen and VanderWeele (2012).

Methods have been developed for inference about these causal effects from observational studies (Tchetgen Tchetgen and VanderWeele 2012, Perez-Heydrich et al. 2014, Liu et al. 2016). One drawback of the treatment effects targeted by these methods is that these causal estimands describe counterfactual scenarios in which individuals

select treatment independently and with the same probability. However, in settings where interference within clusters is plausible, it may be unlikely that treatment selection among individuals in the same cluster is independent (Liu et al. 2016). For instance, suppose a public health policy-maker is interested in the effect of seasonal influenza vaccination on risk of influenza-like illness in households. In this case, one might expect positive correlation between the vaccination statuses of individuals in the same household. Thus, drawing inference to a counterfactual scenario in which individuals are administered vaccines independently may not be of public health relevance. In this paper new causal estimands are proposed for observational studies where there may be clustered interference; these estimands describe counterfactual scenarios in which the treatment selection correlation structure is the same as that in the observed data distribution. By considering scenarios that exhibit within-cluster dependence in the individual treatment selections, the proposed estimands may be more relevant for policy-makers or public health officials who are interested in quantifying the effect of increasing the proportion of treated individuals in a population.

The outline of the remainder of this paper is as follows. In Section 4.2 the potential outcomes framework and interference are discussed. The proposed causal estimands are introduced in Section 4.3. A set of assumptions sufficient for identifying the target estimands is presented in Section 4.4. In Section 4.5 inverse probability-weighted estimators are introduced. The estimators are shown in Section 4.6 to be consistent and asymptotically Normal. Simulations in Section 4.7 demonstrate that the proposed estimators are empirically unbiased and that Wald-type confidence intervals attain nominal coverage levels in finite samples. The proposed methods are illustrated in Section 4.8 by analyzing data from a study of cholera vaccination in over 100,000 individuals in Matlab, Bangladesh. Section 4.9 concludes with a discussion.

4.2 Counterfactuals and Interference

Consider a super-population of clusters of individuals. For each cluster let: N be the number of individuals in the cluster, $A = (A_1, A_2, \ldots, A_N)$ where A_j denotes the binary treatment indicator for individual j in the cluster, and $Y = (Y_1, Y_2, \ldots, Y_N)$ where Y_j is the outcome of interest for individual j. For example, Y_j might indicate whether or not individual j experienced the outcome after some suitable follow-up period after treatment exposure status was observed.

Assuming clustered interference, the potential outcome for an individual may depend on the individual's own treatment exposure status as well as on the treatment exposures of others in the same cluster. However, any individual's potential outcomes are assumed to be unaffected by the treatment exposures of individuals in different clusters. Let $\mathcal{A}(N)$ be the set of all vectors with N binary entries such that $a = (a_1, a_2, \ldots, a_N) \in \mathcal{A}(N)$ is a vector whose entries each indicates a potential treatment status for an individual in a cluster of N individuals. Let $Y_j(a)$ be the potential outcome for unit j in the cluster if, possibly counter to fact, the cluster had received $a \in \mathcal{A}(N)$. In the absence of interference, $Y_j(a) = Y_j(a')$ whenever $a_j = a'_j$ for $a, a' \in \mathcal{A}(N)$. However, assuming no interference when interference is present may result in biased estimates of causal effects. Throughout this paper clustered interference is assumed.

4.3 Causal Effects

4.3.1 Proposed Estimands

Our goal is to draw inference about the difference in expected outcomes arising from population-level policies which change the distribution of treatment. In the absence of interference, typical treatment effect estimands compare the policy (or strategy) where

all individuals receive treatment (i.e., $A = (1, 1, \ldots, 1)$ with probability 1) with the policy where all individuals are not treated (i.e., $A = (0, 0, \ldots, 0)$ with probability 1). Here we consider more general policies where individuals receive treatment according to some probability. Muñoz and van der Laan (2012) refer to such policies as "stochastic interventions." For example, we might consider the policy where individuals select treatment with probability 0.5. In general, let α denote the policy under which the probability an individual is treated equals α, for $\alpha \in [0, 1]$. That is, $\Pr_\alpha(A_j = 1) = \alpha$, where the subscript in $\Pr_\alpha(\cdot)$ indicates that the probability is with respect to the counterfactual scenario in which the policy α is implemented.

For $a \in \mathcal{A}(N)$, define $\omega(a, N, \alpha) = \Pr_\alpha(A = a|N)$ to be the marginal probability under policy α that a cluster of N individuals experiences treatment status a, and let $\bar{\mathrm{Y}}(a) = N^{-1} \sum_{j=1}^{N} Y_j(a)$ denote the average potential outcome in a cluster if the cluster had been exposed to a. The expected potential outcome under α for a single cluster of N individuals is defined to be $\bar{\mathrm{Y}}(\alpha) = \sum_{a \in \mathcal{A}(N)} \bar{\mathrm{Y}}(a)\omega(a, N, \alpha)$. In other words, $\bar{\mathrm{Y}}(\alpha)$ is the expected average potential outcome for the cluster in the counterfactual scenario in which α is implemented.

Define the population mean outcome under α to be $\mu(\alpha) = \mathbb{E}\{\bar{\mathrm{Y}}(\alpha)\}$, where the expected value is taken over all clusters in the super-population. The overall effect is defined to be $\mathrm{OE}(\alpha, \alpha') = \mu(\alpha) - \mu(\alpha')$, which represents the difference in expected potential outcomes under policy α versus policy α'. The overall effect is defined here as a difference in mean potential outcomes, but could instead be defined as a ratio or some other contrast (Liu et al. 2016).

In addition, it may also be of interest to consider potential outcomes among only the untreated individuals within a cluster. Let $\bar{\mathrm{Y}}_t(a) = \{\sum_{j=1}^{N} I(a_j = t)\}^{-1} \sum_{j=1}^{N} Y_j(a) I(a_j = t)$ for $t = 0, 1$. In words, $\bar{\mathrm{Y}}_0(a)$ is the average potential outcome among the untreated individuals within the cluster; likewise $\bar{\mathrm{Y}}_1(a)$ is the average potential outcome among the

treated individuals within the cluster. In the special case when $a = (1-t, 1-t, \ldots, 1-t)$, define $\bar{\mathrm{Y}}_t(a) = 0$ for each of $t = 0, 1$. Denote the population mean potential outcomes when untreated to be $\mu_0(\alpha) = \mathbb{E}\{\sum_{a \in \mathcal{A}(N)} \bar{\mathrm{Y}}_0(a)\omega(a, N, \alpha)\}$. The spillover effect when untreated is defined to be the difference in population mean potential outcomes when untreated under policy α versus α', i.e., $\mathrm{SE}_0(\alpha, \alpha') = \mu_0(\alpha) - \mu_0(\alpha')$. Similarly, let $\mu_1(\alpha) = \mathbb{E}\{\sum_{a \in \mathcal{A}(N)} \bar{\mathrm{Y}}_1(a)\omega(a, N, \alpha)\}$, and define $\mathrm{SE}_1(\alpha, \alpha') = \mu_1(\alpha) - \mu_1(\alpha')$ to be the spillover effect when treated.

Below in Section 4.5, methods are considered for drawing inference about the target estimands for different policies α and α'.

4.3.2 Relation to Existing Estimands

Consider a policy in which all individuals in a cluster are exposed to treatment independently with the same probability; Tchetgen Tchetgen and VanderWeele (2012) refer to this as a "type B parameterisation." For $\alpha \in [0, 1]$, let $\omega_{\mathrm{B}}(a, N, \alpha) = \prod_{j=1}^{N} \alpha^{a_j}(1 - \alpha)^{1-a_j}$ denote the counterfactual probabilities under such a type B policy. Likewise, let $\mu_{\mathrm{B}}(\alpha) = \mathbb{E}\{\sum_{a \in \mathcal{A}(N)} \bar{\mathrm{Y}}(a)\, \omega_{\mathrm{B}}(a, N, \alpha)\}$ be the population mean potential outcome for a type B policy, and define the overall effect with respect to two type B policies to be $\mathrm{OE}_{\mathrm{B}}(\alpha, \alpha') = \mu_{\mathrm{B}}(\alpha) - \mu_{\mathrm{B}}(\alpha')$.

The policies of interest in this paper include as a special case type B policies where treatment exposure is uncorrelated. The estimands proposed in this paper can thus be seen as a generalization of the type B estimands, as the type B policies describe only the limiting counterfactual scenarios in which there is no within-cluster dependence of individual treatment selections. In general, $\omega(a, n, \alpha) \neq \omega_{\mathrm{B}}(a, n, \alpha)$ and the corresponding policies, estimands, and interpretations differ. In the data analysis of the cholera vaccine study in Section 4.8, estimates of the type B estimands are presented for comparison to the estimates of the proposed estimands.

4.4 Identifiability

The counterfactual probabilities $\omega(a, n, \alpha)$ are not identifiable without additional assumptions. Below we assume no unmeasured confounders and parametric models of the conditional distribution of treatment given covariates.

Let there be a random sample of $i = 1, \ldots, M$ clusters, and denote by $O_i = \{N_i, L_i, A_i, Y_i\}$ the observed values of the random variables for cluster i, where L_i is a N_i-vector of baseline (i.e., pre-treatment) variables. The ordering of individuals in each cluster is assumed to be uninformative, and the subscript i is dropped for notational simplicity when not needed. Assume exchangeability conditional on the baseline variables at the cluster level:

$$Y(a) \perp A \mid L, N \text{ for any } a \in \mathcal{A}(N).$$

In addition assume positivity at the cluster level:

$$\Pr(A = a \mid L = l, N = n) > 0 \text{ for all } l, n, \text{ such that}$$
$$\Pr(L = l, N = n) > 0 \text{ and any } a \in \mathcal{A}(n).$$

Following Tchetgen Tchetgen and VanderWeele (2012), Perez-Heydrich et al. (2014), and Liu et al. (2016), assume the following mixed effects logistic regression model for treatment:

$$\Pr(A = a \mid L, N) = \int \prod_{j=1}^{N} \mathcal{L}^{-1}(\beta_0 + \beta_1 L_j + b)^{a_j} \left\{1 - \mathcal{L}^{-1}(\beta_0 + \beta_1 L_j + b)\right\}^{(1-a_j)} d\Phi(b; \sigma), \tag{4.40}$$

where $\mathcal{L}^{-1}(x) = \{1 + \exp(-x)\}^{-1}$ is the inverse-logit function, and b denotes a random intercept for cluster which is assumed to follow a Normal distribution with mean zero, standard deviation σ, and distribution function $\Phi(\cdot)$. We refer to $\Pr(A = a|L, N)$ as a

cluster propensity score (Rosenbaum and Rubin 1983b). These conditional probabilities describe the relationship between observed treatment and covariates; unlike in the case where no interference is assumed, each cluster propensity score is the joint probability of the N individual treatment exposures given covariates.

In addition, assume under counterfactual policy α that

$$\Pr_\alpha(A = a \mid L, N) = \int \prod_{j=1}^{N} \mathcal{L}^{-1}(\gamma_{0\alpha} + \gamma_{1\alpha} L_j + b)^{a_j} \left\{1 - \mathcal{L}^{-1}(\gamma_{0\alpha} + \gamma_{1\alpha} L_j + b)\right\}^{(1-a_j)} d\Phi(b; \phi_\alpha),$$

where the random intercept follows a Normal distribution with mean zero and standard deviation ϕ_α. The model parameters in the counterfactual scenario in general may differ from the parameters in the factual scenario. We similarly refer to $\Pr_\alpha(A = a|L, N)$ as a counterfactual cluster propensity score, as these conditional probabilities describe the relationship between treatment and covariates in the counterfactual scenario in which α is implemented.

The parameters $(\beta_0, \beta_1, \sigma)$ in (4.40) are identifiable from the observable random variables. However, the parameters $(\gamma_{0\alpha}, \gamma_{1\alpha}, \phi_\alpha)$, counterfactual cluster propensity scores $\Pr_\alpha(A = a|L, N)$, and counterfactual probabilities $\omega(a, n, \alpha)$ are not identifiable without additional assumptions. It is assumed here that $\Pr(L) = \Pr_\alpha(L)$, i.e., the different policies do not affect the covariate distribution. Also assume that $\sigma = \phi_\alpha$, i.e., the parameter governing correlation is not affected by different policies. Additionally assume $\beta_1 = \gamma_{1\alpha}$, which supposes that the conditional odds ratio of treatment for any two individuals within the same cluster is the same across the factual and counterfactual scenarios. Under the above assumptions,

$$\alpha = \int \left\{ N^{-1} \sum_{j=1}^{N} \int \mathcal{L}^{-1}(\gamma_{0\alpha} + \beta_1 L_j + b) d\Phi(b; \sigma) \right\} dF_L, \tag{4.41}$$

so the counterfactual model's intercept parameter $\gamma_{0\alpha}$ and thus the counterfactual cluster propensity scores are identifiable. It follows that the counterfactual probabilities $\omega(a, n, \alpha) = \mathbb{E}\{\Pr_\alpha(A = a|L, N = n)\}$ are also identifiable from the observable data.

4.5 Inference

Following Tchetgen Tchetgen and VanderWeele (2012) and Perez-Heydrich et al. (2014), consider the following inverse probability-weighted (IPW) estimator of $\mu(\alpha)$:

$$\hat{\mu}(\alpha) = M^{-1} \sum_{i=1}^{M} \frac{\bar{\mathrm{Y}}_i \, \omega(A_i, N_i, \alpha)}{\Pr(A_i|L_i, N_i)}, \tag{4.42}$$

where $\bar{\mathrm{Y}}_i = N_i^{-1} \sum_{j=1}^{N_i} Y_{ij}$. The inverse probability-weight for cluster i is the reciprocal of the cluster propensity score; these and the counterfactual probabilities are unknown in an observational study and must be estimated from data.

Under the assumptions in Section 4.4, a logistic mixed effects model is fit to the data, and the model parameters $(\beta_0, \beta_1, \sigma)$ can be estimated by maximum likelihood. Then, the fitted parameters $(\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma})$ are substituted into (4.40) to obtain an estimate of each cluster's propensity score. For each policy α, $\hat{\gamma}_{0\alpha}$ solves equation (4.41), with F_L replaced by its empirical distribution; that is, $\alpha = M^{-1} \sum_{i=1}^{M} N_i^{-1} \sum_{j=1}^{N_i} \int \mathcal{L}^{-1}(\gamma_{0\alpha} + \hat{\beta}_1 L_{ij} + b_i) d\Phi(b_i; \hat{\sigma})$ is solved to obtain $\hat{\gamma}_{0\alpha}$. The counterfactual cluster propensity scores for cluster i and treatments $a \in \mathcal{A}(N_i)$ are estimated by substitution, e.g.,

$$\hat{\Pr}_\alpha(A_i = a \mid L_i, N_i) = \int \prod_{j=1}^{N_i} \mathcal{L}^{-1}(\hat{\gamma}_{0\alpha} + \hat{\beta}_1 L_{ij} + b_i)^{a_j} \left\{1 - \mathcal{L}^{-1}(\hat{\gamma}_{0\alpha} + \hat{\beta}_1 L_{ij} + b_i)\right\}^{(1-a_j)} d\Phi(b_i; \hat{\sigma}).$$

Since the ordering of individuals in clusters is assumed to be uninformative, $\omega(a, n, \alpha) = \omega(a', n, \alpha)$ whenever $f(a) = f(a')$ for any two $a, a' \in \mathcal{A}(n)$ where $f(a) = \sum_{j=1}^{n} a_j$.

Let $\mathcal{A}(n,s) = \{a \in \mathcal{A}(n) \,|\, f(a) = s\}$ where $|\mathcal{A}(n,s)| = \binom{n}{s}$, and define $\omega(s,n,\alpha) = \sum_{a\in\mathcal{A}(n,s)} \omega(a,n,\alpha)$ for $s = 0,1,\ldots,n$. Estimate the counterfactual probabilities for any cluster i by $\widehat{\omega}(A_i,N_i,\alpha) = \binom{N_i}{f(A_i)}^{-1}\widehat{\omega}(f(A_i),N_i,\alpha)$, where for any triplet (s,n,α),

$$\hat{\omega}(s,n,\alpha) \;=\; \left\{\sum_{i=1}^{M} I(N_i = n)\right\}^{-1} \sum_{a\in\mathcal{A}(n,s)} \sum_{i=1}^{M} \hat{\Pr}_\alpha(A_i = a|L_i,N_i)I(N_i = n).$$

These estimates, along with the estimated cluster propensity scores, are substituted into (4.42) to calculate $\hat{\mu}(\alpha)$. The estimators $\hat{\mathrm{OE}}(\alpha,\alpha') = \hat{\mu}(\alpha) - \hat{\mu}(\alpha')$ can be obtained in a similar manner. For $t = 0,1$, the estimators $\hat{\mu}_t(\alpha)$ and $\hat{\mathrm{SE}}_t(\alpha,\alpha')$ are defined similarly using the outcomes $\bar{Y}_{t,i} = \{\sum_{j=1}^{N_i} I(A_{ij} = t)\}^{-1}\sum_{j=1}^{N_i} Y_{ij}I(A_{ij} = t)$, where $\bar{Y}_{t,i} = 0$ in the case when $A_{ij} = 1-t$ for all $j = 1,\ldots,N_i$.

In Section 4.6, these estimators are shown to be consistent and asymptotically Normal using standard large-sample estimating equation theory (Stefanski and Boos 2002). Wald-type confidence intervals (CIs) can be constructed using the empirical sandwich estimators of the asymptotic variances.

The estimators described above may be computationally challenging in practice as the estimator $\widehat{\omega}(a,n,\alpha)$ requires a numerical integration technique for each of the $\binom{n}{f(a)}$-many vectors in $\mathcal{A}(n,f(a))$. Therefore, the following approximation is proposed to decrease computation time. For each $s = 0,1,\ldots,n$, define $\mathcal{A}(n,s,k)$ to be a subset of exactly $k_{s,n} = \min\{k, \binom{n}{s}\}$ vectors selected in a simple random sample from $\mathcal{A}(n,s)$, where $k > 1$ is chosen by the investigator. Now estimate the counterfactual probabilities

by $\widehat{\omega}(a,n,\alpha,k) = \binom{n}{f(a)}^{-1}\widehat{\omega}(f(a),n,\alpha,k)$, where for any triplet (s,n,α),

$$\hat{\omega}(s,n,\alpha,k) = \left\{\sum_{i=1}^{M} I(N_i = n)\right\}^{-1} k_{s,n}^{-1}\binom{n}{s} \sum_{a\in\mathcal{A}(n,s,k)} \sum_{i=1}^{M} \hat{\Pr}_\alpha(A_i = a|L_i,N_i)I(N_i = n).$$

Replacing $\widehat{\omega}(a,n,\alpha)$ in $\hat{\mu}(\alpha)$ with $\widehat{\omega}(a,n,\alpha,k)$ results in an estimator which we denote $\hat{\mu}(\alpha,k)$. With analogous replacements define $\hat{\mathrm{OE}}(\alpha,\alpha',k)$, as well as $\hat{\mu}_t(\alpha,k)$ and $\hat{\mathrm{SE}}_t(\alpha,\alpha',k)$ for $t = 0,1$. These estimators are evaluated in a simulation study in Section 4.7 and are employed in the data analysis of the cholera vaccine study in Section 4.8. All of the above estimators are implemented in the R package `clusteredinterference` (Barkley 2018), available on CRAN. In practice, specification of the value of k may be a compromise between less approximation (larger k) and faster computation (smaller k). This method may be extended by specifying different values of k to estimate distinct counterfactual probabilities, which is outlined in Section 4.6. A short discussion on estimating counterfactual probabilities under the assumption of uninformative ordering of individuals within clusters is additionally provided in Section C.1.

4.6 Estimating Equations

The IPW estimators introduced in Section 4.5 are shown to be consistent and asymptotically Normal using standard large-sample estimating equation theory or "M-estimation" (Stefanski and Boos 2002). Presented for illustration below is a simple example where each cluster has exactly n individuals, and at least one cluster $i \leq M$ is observed to experience treatment $f(A_i) = s$ for each $s = 0,1,\ldots,n$. Let $\omega_\alpha = (\omega(0,n,\alpha),\ldots,\omega(n-1,n,\alpha))$ be the ordered vector of the possibly unique counterfactual probabilities excepting $\omega(n,n,\alpha)$; the law of total probability implies that

$\omega(n,n,\alpha) = 1 - \sum_{s=0}^{n-1} \omega(s,n,\alpha)$. Let $\theta_\alpha = (\beta_0, \beta_1, \sigma, \gamma_{0\alpha}, \omega_\alpha, \mu(\alpha))$ be the ordered vector of all parameters to estimate. Next, estimating functions corresponding to each element of θ_α are introduced.

Estimating functions for the parameters $\nu = (\beta_0, \beta_1, \sigma)$ in the logistic mixed model are the score functions of the log likelihood. Let $\psi_\nu = (\psi_{\beta_0}, \psi_{\beta_1}, \psi_\sigma)^\intercal$ be a column vector denoting these estimating functions. For β_1, the estimating function is

$$\psi_{\beta_1}(O_i;\theta_\alpha) = \frac{\partial}{\partial \beta_1} \log\left\{\Pr(A_i|L_i,N_i)\right\},$$

where $\Pr(A_i|L_i,N_i)$ is given in (4.40).For $\gamma_{0\alpha}$, define the estimating function

$$\psi_{\gamma_{0\alpha}}(O_i;\theta_\alpha) = \left\{ N_i^{-1} \sum_{j=1}^{N_i} \int \mathcal{L}^{-1}(\gamma_{0\alpha} + \beta_1 L_{ij} + b_i) d\Phi(b_i;\sigma) \right\} - \alpha.$$

For each $\omega(s,n,\alpha) \in \omega_\alpha$, define the estimating function

$$\psi_{\omega(s,n,\alpha)}(O_i;\theta_\alpha) = \left\{ \sum_{a \in \mathcal{A}(n,s)} \Pr_\alpha(A_i = a|L_i,N_i) - \omega(s,n,\alpha) \right\} I(N_i = n),$$

and let $\psi_{\omega_\alpha} = (\psi_{\omega(0,n,\alpha)}, \psi_{\omega(1,n,\alpha)}, \ldots, \psi_{\omega(n-1,n,\alpha)})^\intercal$. For the target estimand, define

$$\psi_{\mu(\alpha)}(O_i;\theta_\alpha) = \frac{\bar{\mathrm{Y}}_i\, \omega(A_i,N_i,\alpha)}{\Pr(A_i|L_i,N_i)} - \mu(\alpha),$$

where $\omega(A_i,N_i,\alpha) = \binom{N_i}{f(A_i)}^{-1} \omega(f(A_i),N_i,\alpha)$ and where $\Pr(A_i|L_i,N_i)$ is the propensity score for the cluster as in (4.40).

Let $\psi_{\theta_\alpha} = (\psi_\nu, \psi_{\gamma_{0\alpha}}, \psi_{\omega_\alpha}, \psi_{\mu(\alpha)})^\intercal$, and let $q = |\theta_\alpha|$ be the number of parameters to estimate. The estimator $\hat{\theta}_\alpha$ can be expressed as a solution to the following system of

estimating equations:

$$\sum_{i=1}^{M} \psi_{\theta_\alpha}(O_i;\theta_\alpha) = \sum_{i=1}^{M} \begin{bmatrix} \psi_\nu(O_i;\theta_\alpha) \\ \psi_{\gamma_{0\alpha}}(O_i;\theta_\alpha) \\ \psi_{\omega_\alpha}(O_i;\theta_\alpha) \\ \psi_{\mu(\alpha)}(O_i;\theta_\alpha) \end{bmatrix} = 0_{q\times 1}.$$

To show that $\mu(\alpha)$ is the solution to $\int \psi_{\mu(\alpha)}(O|\theta_\alpha)dF_O(O) = 0$, note

$$\begin{aligned} \int \psi_{\mu(\alpha)}(O|\theta_\alpha)dF_O(O) &= \mathbb{E}\left\{ \sum_{a\in\mathcal{A}(N)} \frac{\bar{\mathrm{Y}}(a)\omega(a,N,\alpha)}{\Pr(A=a|L,N)} I(A=a) \right\} \\ &= \mathbb{E}_{L,N}\left[\sum_{a\in\mathcal{A}(N)} \left\{ \mathbb{E}_{A,\{Y(a)\}|L,N}\left(\bar{\mathrm{Y}}(a)\omega(a,N,\alpha) \right) \times \right.\right. \\ &\qquad \left.\left. \mathbb{E}_{A,\{Y(a)\}|L,N}\left(\frac{I(A=a)}{\Pr(A=a|L,N)} \right) \right\} \right] \\ &= \mathbb{E}\left\{ \sum_{a\in\mathcal{A}(N)} \bar{\mathrm{Y}}(a)\omega(a,N,\alpha) \right\}, \end{aligned}$$

which equals $\mu(\alpha)$ by definition, and so $\mu(\alpha)$ solves $\int \psi_{\mu(\alpha)}(O;\theta_\alpha)dF_O(O)$=0. Since ψ_ν are simply the score functions, $\int \psi_\nu(O;\theta_\alpha)dF_O(O) = 0_{|\nu|\times 1}$. Note that the right side of (4.41) equals $\alpha + \int \psi_{\gamma_{0\alpha}}(O;\theta_\alpha)dF_O(O)$, so $\gamma_{0\alpha}$ solves $\int \psi_{\gamma_{0\alpha}}(O;\theta_\alpha)dF_O(O) = 0$. Finally, $\int \psi_{\omega(s,n,\alpha)}(O;\theta_\alpha)dF_O(O) = 0$ follows from $\omega(a,n,\alpha) = \mathbb{E}_L\{\Pr_\alpha(A=a|L,N=n)\}$. Combining these results shows that $\int \psi_{\theta_\alpha}(O;\theta_\alpha)dF_O(O) = 0_{q\times 1}$.

From Stefanski and Boos (2002), $\hat{\theta}_\alpha \xrightarrow{p} \theta_\alpha$ and $\sqrt{M}(\hat{\theta}_\alpha - \theta_\alpha) \xrightarrow{d} N(0,\Sigma_\alpha)$, where $\Sigma_\alpha = U_\alpha^{-1} W_\alpha (U_\alpha^{-1})^\intercal$ for $U_\alpha = \mathbb{E}\{-\dot{\psi}_{\theta_\alpha}(O;\theta_\alpha)\}$ and $W_\alpha = \mathbb{E}\{\psi_{\theta_\alpha}(O;\theta_\alpha)^{\otimes 2}\}$. Consistent estimators for U_α and W_α are $\hat{U}_\alpha = M^{-1}\sum_{i=1}^{M}\{-\dot{\psi}_{\theta_\alpha}(O_i;\theta_\alpha)|_{\theta_\alpha=\hat{\theta}_\alpha}\}$ and $\hat{W}_\alpha = M^{-1}\sum_{i=1}^{M}\{\psi_{\theta_\alpha}(O_i;\hat{\theta}_\alpha)^{\otimes 2}\}$. The empirical sandwich variance estimator $\hat{\Sigma}_\alpha = \hat{U}_\alpha^{-1}\hat{W}_\alpha(\hat{U}_\alpha^{-1})^\intercal$ is consistent for Σ_α, and so $\hat{\mathrm{Var}}(\hat{\mu}(\alpha)) = M^{-1}[\hat{\Sigma}_\alpha]_{[q,q]}$ approximates the variance of $\hat{\mu}(\alpha)$ for large M, where $[\hat{\Sigma}_\alpha]_{[q,q]}$ is the bottom-right element of $\hat{\Sigma}_\alpha$.

An analogous approach is described for $\hat{\text{OE}}(\alpha, \alpha', k)$, where it is now necessary to estimate $\gamma_{0\alpha'}$ and $\omega_{\alpha'}$ as well. Let $\theta_{\alpha,\alpha'} = (\nu, \gamma_{0\alpha}, \gamma_{0\alpha'}, \omega_\alpha, \omega_{\alpha'}, \text{OE}(\alpha, \alpha'))$ be the ordered vector of all parameters to estimate. For each $\omega(s, n, \alpha) \in \omega_\alpha$, define the estimating function

$$\psi_{k,\omega(s,n,\alpha)}(O_i; \theta_{\alpha,\alpha'}) \quad = \\ \left\{ k_{s,n}^{-1} \binom{n}{s} \sum_{a \in \mathcal{A}(n,s,k)} \Pr_\alpha(A_i = a | L_i, N_i) - \omega(s, n, \alpha) \right\} I(N_i = n),$$

and let $\psi_{k,\omega_\alpha} = (\psi_{k,\omega(0,n,\alpha)}, \psi_{k,\omega(1,n,\alpha)}, \ldots, \psi_{k,\omega(n-1,n,\alpha)})^\intercal$. For the target estimand, define

$$\psi_{\text{OE}(\alpha,\alpha')}(O_i; \theta_{\alpha,\alpha'}) = \frac{\bar{\text{Y}}_i \{\omega(A_i, N_i, \alpha) - \omega(A_i, N_i, \alpha')\}}{\Pr(A_i | L_i, N_i)} - \text{OE}(\alpha, \alpha').$$

It is easily shown that $\int \psi_{\text{OE}(\alpha,\alpha')}(O; \theta_{\alpha,\alpha'}) dF_O(O) = 0$ using a proof analogous to the one for $\psi_{\mu(\alpha)}$ presented above. In a similar manner, that $\int \psi_{k,\omega(s,n,\alpha)}(O; \theta_{\alpha,\alpha'}) dF_O(O) = 0$ follows directly from $\int \psi_{\omega(s,n,\alpha)}(O; \theta_\alpha) dF_O(O) = 0$. Finally, let $\psi_{k,\theta_{\alpha,\alpha'}} = (\psi_\nu, \psi_{\gamma_{0\alpha}}, \psi_{\gamma_{0\alpha'}}, \psi_{k,\omega_\alpha}, \psi_{k,\omega_{\alpha'}}, \psi_{\text{OE}(\alpha,\alpha')})^\intercal$. Then $\theta_{\alpha,\alpha'}$ solves $\int \psi_{k,\theta_{\alpha,\alpha'}}(O; \theta_{\alpha,\alpha'}) dF_O(O) = 0_{q' \times 1}$ and $\hat{\theta}_{\alpha,\alpha'}$ solves $\sum_{i=1}^M \psi_{k,\theta_{\alpha,\alpha'}}(O_i; \theta_{\alpha,\alpha'}) = 0_{q' \times 1}$ for $q' = |\theta_{\alpha,\alpha'}|$, and the above results follow.

The difference in $\hat{\text{OE}}(\alpha, \alpha')$ and $\hat{\text{OE}}(\alpha, \alpha', k)$ arises solely from the estimating functions used for the counterfactual probabilities, i.e., ψ_{ω_α} and ψ_{k,ω_α}, respectively. When $\binom{n}{\lfloor n/2 \rfloor} \leq k$, then $\mathcal{A}(n, s) = \mathcal{A}(n, s, k)$ for all s, and ψ_{ω_α} is equivalent to ψ_{k,ω_α}. As mentioned above, one could use different values of k for distinct estimating equations. For example, one could estimate $\omega(s, n, \alpha)$ with $\psi_{k,\omega(s,n,\alpha)}$ and $\omega(s', n', \alpha)$ with $\psi_{k',\omega(s',n',\alpha)}$, where $\omega(s, n, \alpha) \neq \omega(s', n', \alpha)$ and $k \neq k'$, and the above results would still apply.

4.7 Simulations

A simulation study was carried out on 1000 datasets to demonstrate the finite-sample performance of the proposed estimators. To generate each dataset, the following steps were carried out for each of $i = 1, \ldots, M = 125$ clusters:

I The number of individuals in the cluster N_i was simulated such that $\Pr(N_i = 8) = 0.4$, $\Pr(N_i = 22) = 0.35$, and $\Pr(N_i = 40) = 0.25$.

II Covariates for each individual $j = 1, \ldots, N_i$ in cluster i were simulated to be $L_{ij1} \sim N(40, 5)$ and $L_{ij2} \sim N(X_i, 0.2)$, where $X_i \sim N(6, 1)$ was a cluster-level random variable.

III Treatment status A_{ij} for each individual j in cluster i was simulated from a Bernoulli distribution with mean $\Pr(A_{ij} = 1|L_{ij}, b_i) = \mathcal{L}^{-1}(\beta_0 + \beta_1 L_{ij1} + \beta_2 L_{ij2} + b_i)$ where $b_i \sim N(0, \sigma)$ was a cluster-level random intercept and $(\beta_0, \beta_1, \beta_2, \sigma) = (0.75, -0.015, -0.025, 0.75)$.

IV The outcome Y_{ij} for each individual j in cluster i was simulated from a Bernoulli distribution with mean $\Pr(Y_{ij} = 1|A_i, L_{ij}) = \mathcal{L}^{-1}(0.1 - 0.05L_{ij1} + 0.5L_{ij2} - 0.5A_{ij} + 0.2g(A_{i,-j}) - 0.25A_{ij}g(A_{i,-j}))$, where the function $g(A_{i,-j}) = (N_i - 1)^{-1} \sum_{j' \neq j} A_{ij'}$.

A logistic mixed effects model was fit with a random intercept for cluster and main effects for L_1 and L_2, i.e., the propensity score models were correctly specified. To determine the performance of the estimators that use the greatest degree of sub-sampling approximation, $k = 1$ was chosen. The asymptotic variance of the estimators was estimated with the empirical sandwich variance estimator as described in Section 4.6, from which Wald-type 95% CIs were constructed.

True values of the estimands for policies $\alpha \in \{0.4, 0.5, 0.55\}$ were determined empirically using the same data generating process outlined above in steps I-II and analogues

to steps III-IV. The process is described here briefly, with more details provided in Section C.2. For each α, $\gamma_{0\alpha}$ was determined by solving (4.41) with F_L approximated by its empirical distribution over 10^7 clusters. Then, the counterfactual probabilities $\omega(a, n, \alpha)$ were determined by generating treatment vectors under policy α for 10^8 clusters, replacing β_0 in step III with $\gamma_{0\alpha}$. An empirical comparison of true values of $\omega(a, n, \alpha)$ arising from this simulation study and the true values of $\omega_{\mathrm{B}}(a, n, \alpha)$ for the type B policies is provided in Figure C.1 in Section C.2.5. Next, potential outcomes were generated for 10^8 clusters via the causal model analogous to the regression model specified in step IV. These potential outcomes were combined with the counterfactual probabilities to determine the true values of $\mu(\alpha)$, $\mathrm{OE}(\alpha, \alpha')$, and $\mu_t(\alpha)$ and $\mathrm{SE}_t(\alpha, \alpha')$ for $t = 0, 1$.

The IPW estimates from each dataset were compared to the true estimand values determined above; a summary of these results is presented in Table 4.6. The average bias of the estimators was negligible. The average of the estimated asymptotic standard errors was approximately equal to the empirical Monte Carlo standard error. The Wald-type 95% CIs contained the true parameter values for approximately 95% of the simulated datasets. Thus, the estimators performed well in this simulation study.

4.8 Data Analysis

The proposed methods are illustrated in the following analysis of a cholera vaccine study in Matlab, Bangladesh, which featured both an experimental and a non-experimental component (Ali et al. 2005; 2009). Included in the study were 121,975 women (aged 15 years and older) and children (aged 2-15 years) from 6,415 baris (i.e., households of patrilineally-related individuals). These individuals were eligible to participate in the experimental component of the study, in which each individual was randomized with equal probability to one of three treatment arms: B subunit-killed

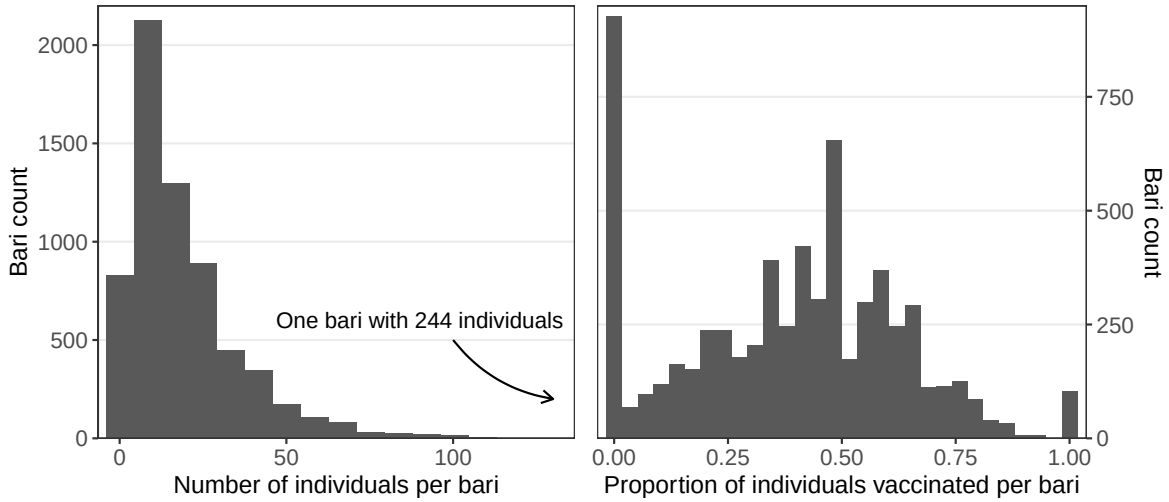
Table 4.6: Summary of results from simulation study described in Section 4.7. Truth denotes the true value of the estimand targeted by the estimator; Bias denotes the average bias of the IPW estimates over the 1000 datasets; Cov% denotes the empirical coverage of Wald-type 95% CIs; ASE denotes the average of the estimated sandwich standard errors times 100; ESE denotes the empirical standard error times 100; SER denotes the ratio of ASE divided by ESE; $\alpha_1 = 0.4$, $\alpha_2 = 0.5$, and $\alpha_3 = 0.55$.

Estimator	Truth	Bias	Cov%	ASE	ESE	SER
$\hat{\mu}(\alpha_1, k=1)$	0.662	-0.003	94.3%	1.88	1.84	1.02
$\hat{\mu}(\alpha_2, k=1)$	0.651	0.000	95.5%	1.63	1.53	1.06
$\hat{\mu}(\alpha_3, k=1)$	0.645	0.001	96.4%	1.65	1.55	1.07
$\hat{\text{OE}}(\alpha_2, \alpha_1, k=1)$	-0.011	0.003	97.2%	1.08	0.96	1.13
$\hat{\text{OE}}(\alpha_3, \alpha_1, k=1)$	-0.017	0.004	97.4%	1.44	1.34	1.08
$\hat{\text{OE}}(\alpha_3, \alpha_2, k=1)$	-0.006	0.001	97.4%	0.53	0.44	1.21
$\hat{\mu}_0(\alpha_1, k=1)$	0.712	-0.002	95.2%	2.10	2.02	1.04
$\hat{\mu}_0(\alpha_2, k=1)$	0.711	-0.001	95.7%	2.15	2.02	1.07
$\hat{\mu}_0(\alpha_3, k=1)$	0.709	-0.001	95.3%	2.46	2.35	1.05
$\hat{\text{SE}}_0(\alpha_2, \alpha_1, k=1)$	-0.001	0.001	95.8%	1.33	1.20	1.11
$\hat{\text{SE}}_0(\alpha_3, \alpha_1, k=1)$	-0.003	0.001	94.7%	1.93	1.86	1.04
$\hat{\text{SE}}_0(\alpha_3, \alpha_2, k=1)$	-0.002	0.000	94.8%	0.79	0.72	1.10
$\hat{\mu}_1(\alpha_1, k=1)$	0.573	0.007	94.2%	3.04	3.09	0.99
$\hat{\mu}_1(\alpha_2, k=1)$	0.581	0.004	95.0%	2.25	2.24	1.01
$\hat{\mu}_1(\alpha_3, k=1)$	0.582	0.001	95.3%	2.10	2.07	1.01
$\hat{\text{SE}}_1(\alpha_2, \alpha_1, k=1)$	0.008	0.003	94.9%	1.51	1.46	1.04
$\hat{\text{SE}}_1(\alpha_3, \alpha_1, k=1)$	0.009	0.005	95.2%	2.02	1.98	1.02
$\hat{\text{SE}}_1(\alpha_3, \alpha_2, k=1)$	0.002	0.002	96.4%	0.65	0.57	1.13

whole cell oral cholera vaccine, killed whole cell-only oral cholera vaccine, or placebo. Individuals who did not participate did not receive either version of active treatment. The study collected endpoint data of cholera infection on all individuals, even those who did not participate in the experimental component. Since participation was not controlled by study design and nearly two-fifths of all individuals declined to participate, there was a notable non-experimental component to the study, and potential for confounding exists when analyzing the endpoint data.

As in Perez-Heydrich et al. (2014), any individual who received at least two doses of either of the two cholera vaccines was considered to be treated, and otherwise was

Figure 4.9: Univariate data summaries from the Matlab cholera vaccine study. Left: number of individuals per cluster (bari). Right: proportion of individuals vaccinated per cluster.



considered to be untreated. Clustered interference was assumed at the level of the bari as there is evidence that transmission of cholera often takes place within baris (Ali et al. 2005). Figure 4.9 illustrates the empirical distributions of the number of individuals and of the treatment coverage within the baris.

Cluster-level conditional exchangeability and positivity were assumed to hold conditional on age and distance from the bari to the nearest river. A logistic mixed effects model was fit, regressing the indicator that an individual obtained treatment on the individual's age and river distance with a random intercept for the bari in which the individual lived. Included in the mixed model were a linear term for distance (in kilometers) and linear and quadratic terms for age (centered, in decades). The variance component of the random intercept was estimated to be $\hat{\sigma} = 0.91$ with 95% CI (0.89, 0.94) calculated via profile likelihood using `confint.merMod()` from the R package `lme4` (Bates et al. 2015), indicating significant correlation between individual treatment statuses within clusters.

The proposed methods were carried out with the logistic mixed effects model described above. All the assumptions for identifiability as discussed in Section 4.4 were assumed. The IPW estimators were computed with $k = 3$, and Wald-type CIs were constructed from the empirical sandwich variance estimator.

Figure 4.10: Estimates of the population mean estimands from the analysis of the Matlab cholera vaccine study. The light green diamonds indicate $\hat{\mu}(\alpha, k = 3)$. The dark blue circles indicate $\hat{\mu}_0(\alpha, k = 3)$, and the light pink squares indicate $\hat{\mu}_1(\alpha, k = 3)$. The dark brown $\times$'s indicate $\hat{\mu}_{\mathrm{B}}(\alpha)$, which target the type B estimands from Tchetgen Tchetgen and VanderWeele (2012). All estimates are multiplied by 1000.

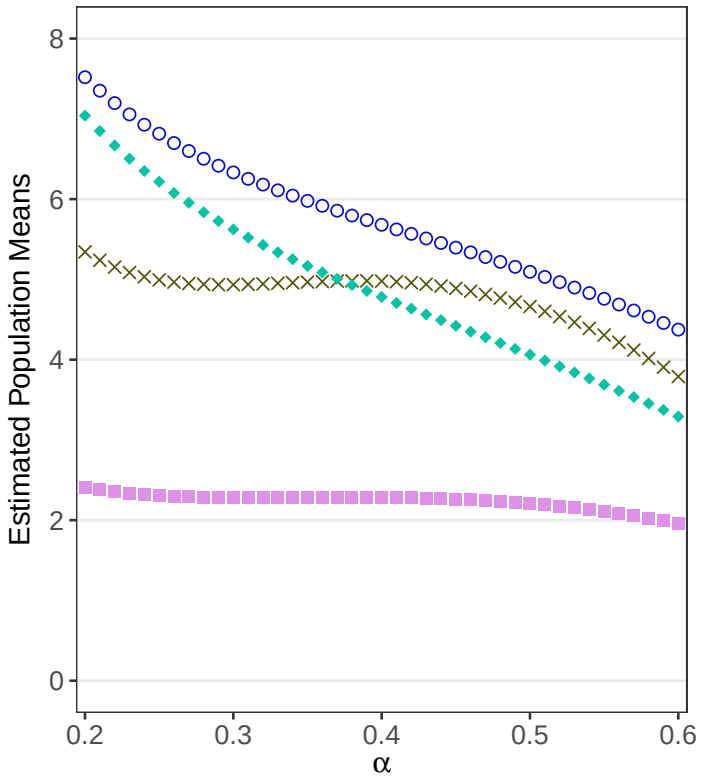
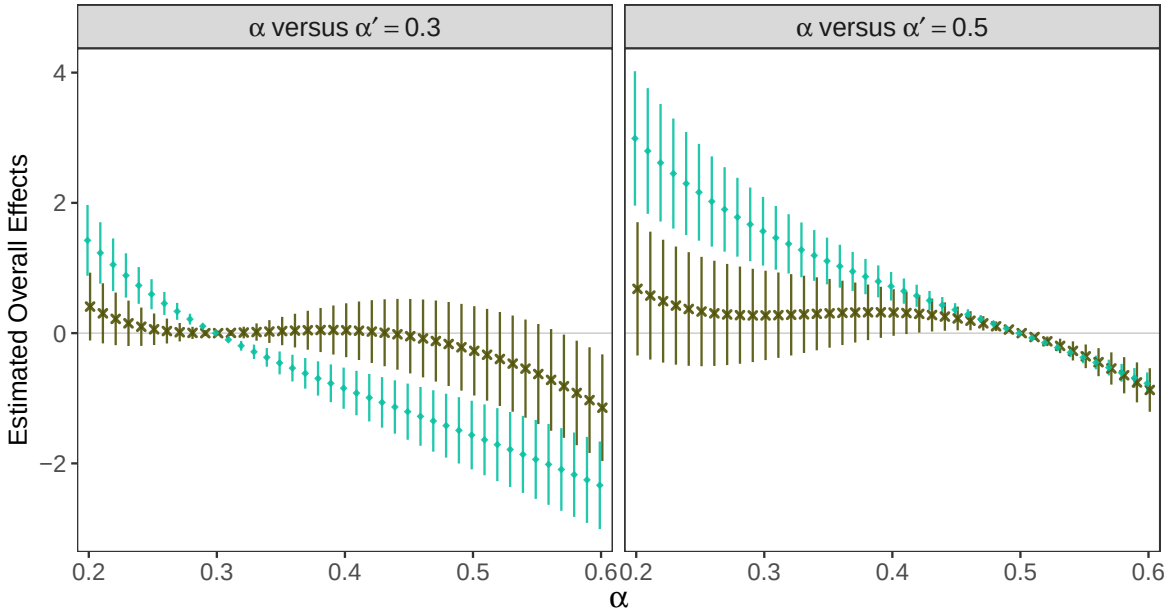


Figure 4.10 depicts point estimates of the population mean estimands over policies ranging from $\alpha = 0.2$ to $\alpha = 0.6$. Estimates are presented in units of one case of cholera infection per 1000 individuals per year. Estimates of $\mu_1(\alpha)$ were relatively invariant to α, suggesting minimal spillover effects when an individual is vaccinated. In contrast, estimates of $\mu_0(\alpha)$ decreased noticeably as α increased, suggesting a protective spillover effect when an individual is not vaccinated. The estimates of $\mu(\alpha)$ similarly suggest

lower risk of cholera infection at the population level for policies with greater levels of vaccine coverage.

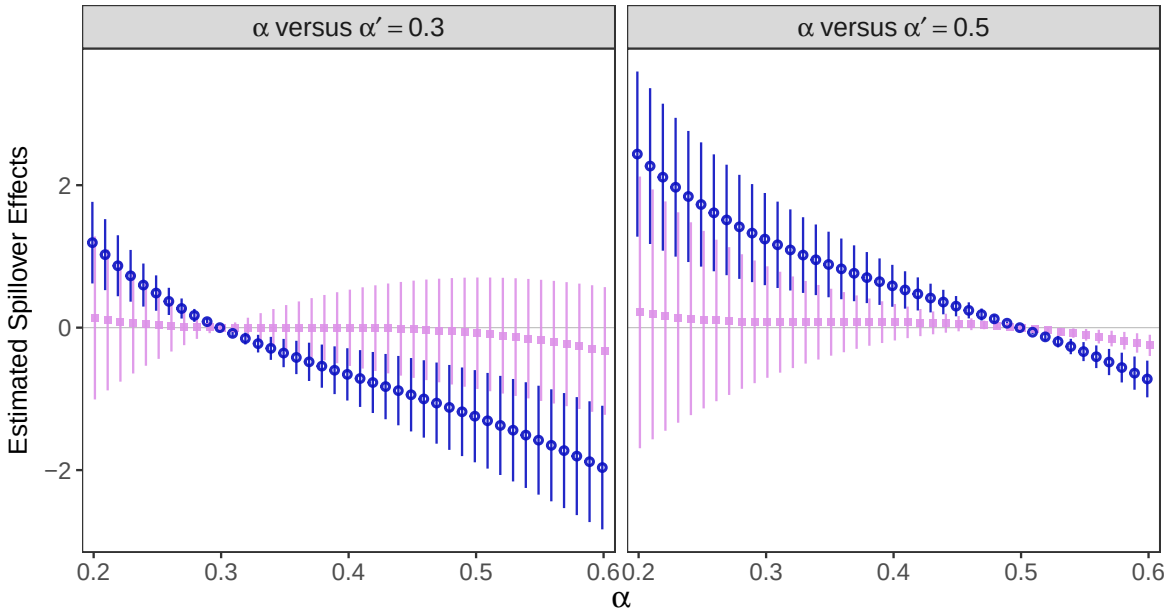
Figure 4.11: Estimated overall effects from the analysis of the Matlab cholera vaccine study for selected contrasts. The diamonds and light green lines indicate the point estimates and 95% CIs from $\hat{\text{OE}}(\alpha, \alpha', k = 3)$. The $\times$'s and dark brown lines indicate the point estimates and 95% CIs from $\hat{\text{OE}}_{\text{B}}(\alpha, \alpha')$, which target the type B estimands from Tchetgen Tchetgen and VanderWeele (2012). All estimates are multiplied by 1000.



Overall effect estimates and corresponding 95% CIs are depicted in Figure 4.11. Negative effects are favorable, corresponding to a reduction in cholera infections. For example, $\hat{\text{OE}}(0.45,\ 0.3, k = 3) = -1.2$ (95% CI $-1.6, -0.8$), indicating a significant protective effect of policy $\alpha = 0.45$ compared to $\alpha = 0.3$. In particular, we expect 1.2 fewer cases of cholera per 1000 person-years if there is 45% vaccine coverage compared to 30% vaccine coverage.

Estimated spillover effects are depicted in Figure 4.12. The estimates of $\hat{\text{SE}}_1(\alpha, \alpha', k = 3)$ were approximately zero and the CIs included zero for almost all contrasts shown, indicating mostly negligible spillover effects among treated individuals within clusters. However, $\hat{\text{SE}}_0(\alpha, \alpha', k = 3)$ was negative for $\alpha > \alpha'$ and positive for $\alpha < \alpha'$, and all

Figure 4.12: Estimated spillover effects from the analysis of the Matlab cholera vaccine study for selected contrasts. The circles and dark blue lines indicate the point estimates and 95% CIs from $\hat{\text{SE}}_0(\alpha, \alpha', k=3)$. The squares and light pink lines indicate the point estimates and 95% CIs from $\hat{\text{SE}}_1(\alpha, \alpha', k=3)$. All estimates are multiplied by 1000.



of the CIs excluded zero. Thus there is evidence of a protective effect of policies with higher probability of treatment exposure conferred to individuals who did not themselves obtain treatment.

Figures 4.10 and 4.11 also depict point estimates of the type B estimands and corresponding 95% CIs, computed using the R package `inferference` (Saul and Hudgens 2017) based on the same logistic mixed effects propensity score model employed with the proposed estimators. Relative to the estimates of the proposed estimands, the estimates of the type B estimands were smaller with corresponding 95% CIs that often included zero. For example, $\hat{\text{OE}}_{\text{B}}(0.2, 0.5) = 0.7$ (95% CI $-0.3, 1.7$), while $\hat{\text{OE}}(0.2, 0.5, k = 3) = 3.0$ (95% CI $2.0, 4.0$). Thus, inferences based on the type B estimands tended to underestimate the population-level utility of cholera vaccination compared to results based on the proposed estimands.

4.9 Discussion

Drawing causal inference from observational data when interference may be present poses several challenges, including defining the causal effects of interest. Proposed in this paper are causal estimands for use in observational studies when clustered interference is plausible. The proposed causal effects are contrasts in mean potential outcomes arising from different policies that change the distribution of treatment. IPW estimators were proposed and shown to be consistent and asymptotically Normal under certain identifying assumptions, and empirical sandwich estimators were derived for the asymptotic variance of the estimators. The IPW estimators performed well in finite samples with minimal bias, and the Wald-type confidence intervals attained nominal coverage levels. These methods were illustrated in an analysis of a large cholera vaccine study, providing evidence that increasing the proportion of individuals vaccinated reduces cholera infections.

The policy effects considered here may be more relevant in public health settings such as infectious disease research because within-cluster characteristics are incorporated into the proposed estimands. To reduce the burden of infectious diseases through vaccination programs, it is important to consider the "ecological circumstances" of the disease (Ali et al. 2009). Previously proposed type B estimands define treatment effects in the counterfactual scenario individuals are independently exposed to treatment. However, scenarios in which treatment exposures are correlated may represent more relevant ecological circumstances. Aside from controlled trials, in general one might expect treatment correlation in settings where interference is present. Indeed, the cholera vaccine study analysis in Section 4.8 indicates strong evidence of treatment correlation within clusters. Unlike the type B estimands, the proposed estimands describe effects of population-level policies where the treatment correlation is the same as in the observed data distribution. Likewise, the proposed estimands preserve the conditional

odds ratio of treatment for any two individuals within the same cluster. By incorporating these within-cluster features, inferences targeting the proposed estimands may be of greater relevance to public health investigators and policy-makers concerned with controlling the spread of infectious disease in a particular population.

Consistency of IPW estimators is in general dependent on the correct specification of the treatment model. The estimators presented here also require that the model for the counterfactual distribution of treatment is correctly specified. Although non-parametric methods might be employed instead to improve robustness to misspecification of the treatment model, such methods may impede identifiability of the target causal estimands without further untestable identifying assumptions. The proposed methods may suffer computational challenges due to a large number of nuisance parameters, depending on the joint distribution of (A, N). Future work may consider reducing the number of nuisance parameters, perhaps through approximating the counterfactual treatment distribution. Future research may also consider assuming different structures of interference to better align with the epidemiology of cholera; e.g., see Ali et al. (2018).

Although this work is motivated by infectious disease research, it is applicable in many other areas in which interference may be present. For example, Papadogeorgou et al. (2017) are currently and independently developing similar estimands and methods with motivation from and applications in air pollution epidemiology. By defining causal effects of population-level interventions (Westreich 2017) in the presence of interference, the proposed estimands may have greater practical utility and be more relevant to investigators and policy-makers.

CHAPTER 5: CONCLUSION

Data analysis is instrumental for innovation in personal and public health. Estimating the causal effect of treatment on health outcomes plays an invaluable role in this endeavor, but can be difficult to carry out when presented with data from observational studies. Additional analytic challenges arise when one individual's outcome may be affected by another individual's treatment. This phenomenon is often called interference, and it is plausible in certain areas such as infectious disease research. In this document we developed statistical methodology for drawing causal inference from observational studies in the presence of partial interference, i.e., when interference may exist within clusters of individuals, but not between distinct clusters. We introduced two new methods for estimating existing causal parameters from an observational study assuming partial interference, each of which was shown to perform better than existing estimators. We also introduced a new set of causal parameters that may be more relevant in this scenario, for which a set of estimators were also proposed.

We introduced a modeling technique for IPW estimation of the estimands in Tchetgen Tchetgen and VanderWeele (2012) from observational studies in the presence of partial interference. This flexible method for estimating cluster propensity scores used data-adaptive modeling assumptions in order to reduce the risk of bias due to model mis-specification. We applied the GMERT algorithm (Hajjem et al. 2017), which combines machine learning and mixed modeling techniques, to a training sample of data in order to determine the relationship between predictors and treatment under a presumed correlation structure. We then used the decision rules that were recovered from

this algorithm to model the relationship between treatment and covariates in the test sample. In a finite-sample simulation study, the proposed methods exhibited less bias than existing methods, and also achieved nominal 95% coverage from the sandwich variance estimator of the asymptotic variance.

A set of estimators based on covariate matching techniques was also introduced for this scenario. Considering each cluster of individuals to be a study unit, we matched these clusters to each other to estimate causal effects in the presence of partial interference. This method extends related research from Abadie and Imbens (2006) and Yang et al. (2016), and does not require explicit modeling assumptions. These estimators were shown to be consistent and asymptotically Normal under certain assumptions. In a finite-sample simulation study the matching estimators exhibited low bias and achieved nominal 95% coverage; they also outperformed existing IPW estimators.

Lastly we proposed a new set of causal estimands for observational studies in the presence of partial interference. In contrast to the estimands in Tchetgen Tchetgen and VanderWeele (2012), the proposed estimands allow for within-cluster dependence in the individual treatment selections. The proposed estimands may have greater relevance or practical utility to public health officials or policy-makers in some scenarios, e.g., determining whether increasing the proportion of treated individuals in a population in a non-experimental manner would result in improved health outcomes. Presented were a sufficient set of assumptions for identifying these estimands from an observational study. IPW estimators were introduced, and were shown to be consistent and asymptotically Normal for the proposed estimands. The IPW estimators were shown to perform well in a finite-sample simulation study, exhibiting low bias and achieving nominal 95% coverage.

Each of the three methods introduced here was illustrated in an analysis of a cholera vaccine study in Matlab, Bangladesh. Although each method relied on a different set of

assumptions, the results from each analysis showed that increased amounts of vaccine treatments reduced the risk of infection by cholera. These results were relatively similar across each of the methods illustrated here, and also similar to related results from existing analyses, which provides further evidence that increased vaccine allocation confers health benefits in a population of individuals.

APPENDIX A: TECHNICAL DETAILS FOR CHAPTER 2

A.1 Setup and Additional Results of Data Analysis

We first present additional details relating to the cross-validation step used in the analysis of the cholera vaccine study. In Section A.1.4 we present additional estimates of target estimands from the GMERT-IPW and LMM-IPW procedures that were not included in the main paper due to space concerns.

A.1.1 Training and Testing Samples

All clusters i with $N_i < 5$ individuals were excluded from the training sample (and instead included in the testing sample). The cluster with $N_i = 244$ individuals was included in the training sample: large clusters perhaps result in very small estimated cluster propensity scores (and large weights), and there may be efficiency to be gained in including larger clusters in the training sample rather than the testing sample (because the learning algorithm can learn on the $N_i = 244$ correlated data observations, whereas the cluster only counts for 1 i.i.d. study unit in standard asymptotic theory). Finally, each remaining cluster was assigned to the training sample with probability 1/4, and to the testing sample with probability 3/4.

Figure A.1 illustrates the number of individuals in per bari, where the baris in the training sample are shown in the bottom panel, and the baris in the testing sample are in the top panel. Noticeably, there are no baris in the lowest category of bari size in the training sample, as we've chosen to assign all baris with fewer than 5 individuals to the testing sample instead. Otherwise, the empirical distribution of bari sizes seems similar across the two samples, providing some evidence that the learning algorithm may generalize well from the training sample to the testing sample.

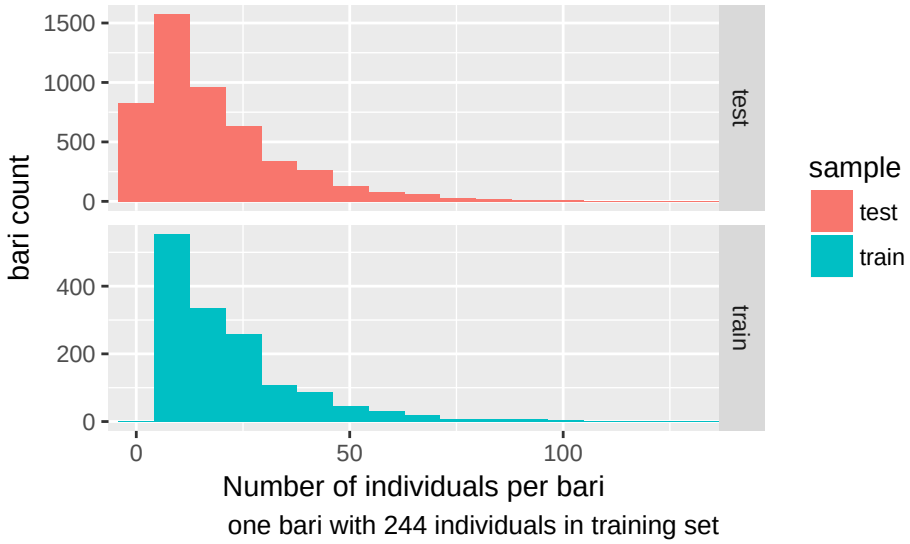


Figure A.1: Number of individuals per bari, by data sample partition.

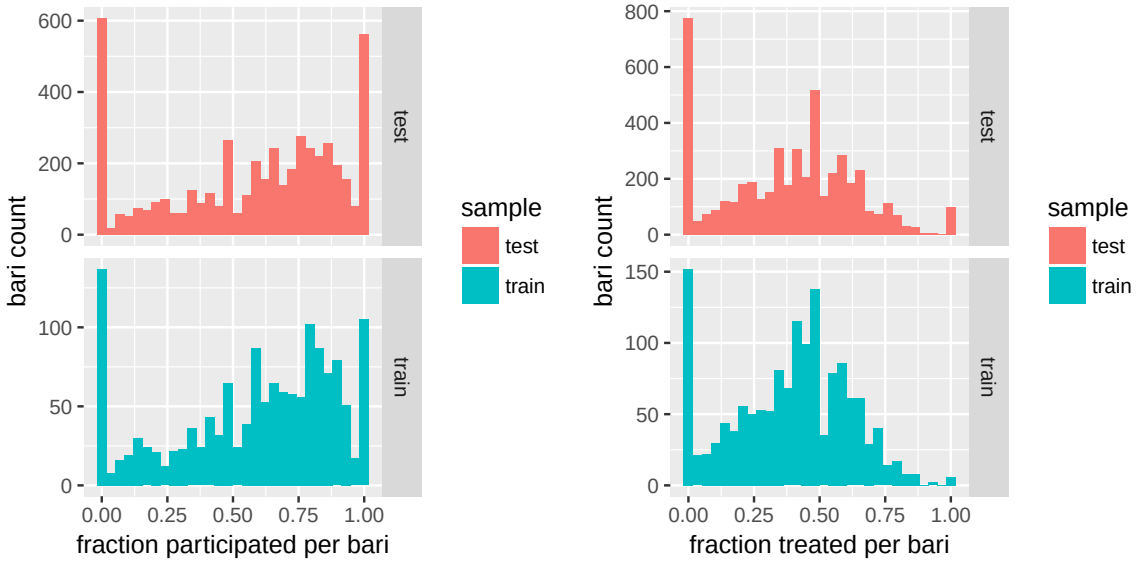


Figure A.2: Summary of propensity to select treatment by data sample partition. The left panels reflect individual participation, and the right panels reflect treatment received.

Figure A.2 illustrates a summary of the propensity of individuals to self-select for treatment. The panels on the left side present a summary of the fraction of individuals within each bari who self-selected to participate in the randomized trial. The panels on the right side present the fraction of individuals who received at least two doses of the active treatment. Again, the baris in $\mathcal{S}_{test}$ are represented in the top panels, and the baris in $\mathcal{S}_{train}$ are represented in the bottom panel. In both cases - participation and treatment - the empirical distribution seems very similar across the two data samples. This provides more evidence that our training sample is satisfactorily similar to the testing sample.

A.1.2 Cross-Validation Folds

In more detail, we chose without replacement K individuals from each cluster $i \in \mathcal{S}_{train}$, and then without replacement we chose exactly one individual from each of those K individuals to be in $\mathcal{S}_{cv,k}$ and assigned the remaining $K-1$ individuals to $\mathcal{S}_{tr,k}$ for each $k = 1, \ldots, K$. This guaranteed that every cluster had at least one individual in $\mathcal{S}_{tr,k}$ and at least one individual $\mathcal{S}_{cv,k}$ for each $k = 1, \ldots, K$. Then each of the $j = 1, \ldots, (N_i - K)$ individuals in each cluster i was randomly assigned a number $k_{ij} \in \{1, \ldots, K\}$ such that the individual was assigned to $\mathcal{S}_{cv,k'}$ when $k' = k_{ij}$ or to $\mathcal{S}_{tr,k'}$ when $k' \neq k_{ij}$. This ensured that every individual in $\mathcal{S}_{train}$ was used in a training sub-sample at least once, and in a cross-validation sub-sample at least once.

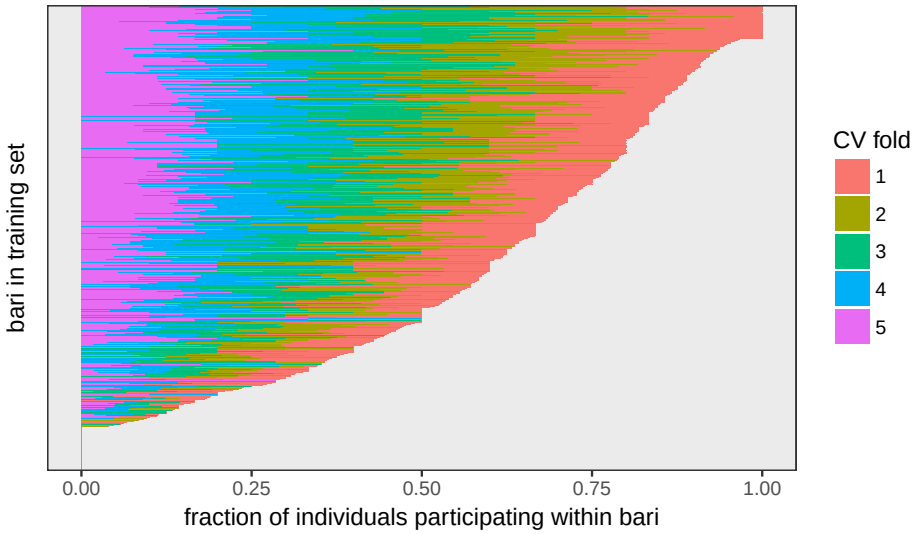


Figure A.3: Distribution of treatment-seeking individuals over the 5 cross validation folds in the training sample

Figure A.3 illustrates a summary of how the cross-validation folds partitioned the individuals seeking treatment. Each bari $i \in \mathcal{S}_{train}$ is represented by one horizontal bar, where the total width of the bar indicates the fraction of individuals who self-selected to participate in the randomized component (i.e., sought treatment). The 5 colors partitioning the bars correspond to the 5 folds of data for cross validation. Each color represents the individuals in that fraction of participants who belong to the k^{th} cross validation fold (i.e., $\mathcal{S}_{cv,k}$). The colors seem to be evenly distributed, indicating that the cross-validation folds we have chosen may be satisfactory for selecting appropriate tuning parameters.

A.1.3 Selecting Tuning Parameters

An individual's predicted response was determined to be $\hat{B} = 1$ if $\hat{\Pr}(B = 1|L) >= 0.5$, and $\hat{B} = 0$ otherwise. Misclassification error for the individual was coded as 1 for an individual if $\hat{B}_{ij} \neq B_{ij}$, and 0 otherwise. Misclassification was weighted by clusters: misclassification errors were averaged for each cluster, and then over clusters.

Table A.1: Sets of tuning parameters considered in the data analysis. Each row indicates one set of tuning parameters, and each column is named for the corresponding argument in `rpart` (Therneau et al. 1997). The `minsplit` column indicates the values of the tuning parameters for minimum observations in a node necessary to split the node. The `maxdepth` column indicates the maximum depth of a terminal node of a tree. The `minbucket` column indicates minimum observations in any terminal node of the tree.

Tuning set	`minsplit`	`maxdepth`	`minbucket`
1	50	5	10
2	25	5	10
3	50	15	10
4	25	15	10
5	50	5	30
6	50	15	30

Another option would be to take the unweighted average misclassification error over all individuals - regardless of cluster - this which may represent preference towards algorithms that perform well on larger baris.

Table A.2: Weighted misclassification error (wMCe) for all proposed sets of tuning parameters across all $K = 5$ CV folds. Each row corresponds to one set of tuning parameters, enumerated as in Table A.1. The column for Mean wMCe indicates that tuning parameter sets 5 has the lowest mean error in the CV sets, which is then selected for the GMERT-IPW procedure.

Tuning set	wMCe_1	wMCe_2	wMCe_3	wMCe_4	wMCe_5	Mean wMCe
1	20.88%	20.72%	20.62%	20.63%	20.74%	20.715%
2	20.88%	20.72%	20.62%	20.74%	20.74%	20.737%
3	20.77%	20.72%	20.65%	20.64%	20.75%	20.706%
4	20.77%	20.72%	20.65%	20.64%	20.75%	20.704%
5	20.90%	20.67%	20.63%	20.57%	20.70%	20.695%
6	20.74%	20.72%	20.63%	20.64%	20.79%	20.703%

A.1.4 Results

Additional results from the GMERT-IPW and LMM-IPW techniques are presented here. For the logistic mixed model used in the GMERT-IPW method, Figure A.4 presents a summary of how the estimated fixed effects (from the testing sample) different

from their predicted values (on the training sample). There is perhaps mild evidence of the algorithm overfitting the training sample, as indicated by the more extreme values of some of the nodes at the tails.

Figure A.4: Estimated fixed effects for terminal nodes from GMERT algorithm, in training and testing samples from the cholera vaccine study. The nodes are ordered by increasing probability in the training sample. There is a suggestion of mild overfitting in the training sample, indicated by the more extreme probabilities for the training sample at the tails.

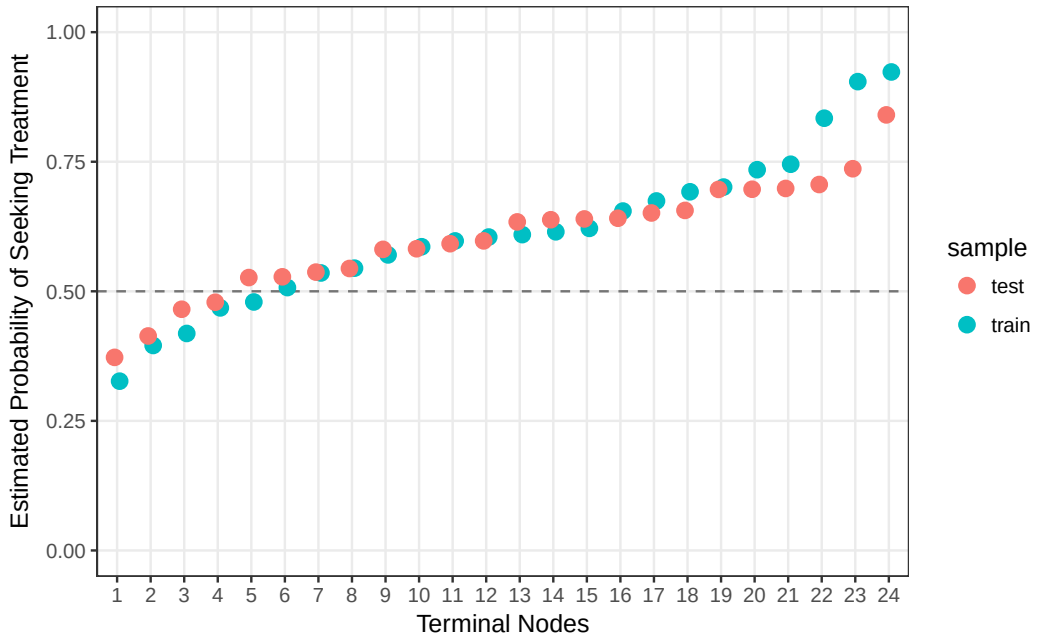


Figure A.5 presents the estimated values of $\mu(1, \alpha)$ and $\mu(0, \alpha)$ for various levels of α. Figure A.6 presents estimates of the Total and Indirect effects.

Figure A.5: Estimates and 95% confidence intervals for population mean estimands $\mu(z, \alpha)$ where $z = 0, 1$ and $\alpha \in \{0.3, 0.45, 0.6\}$.

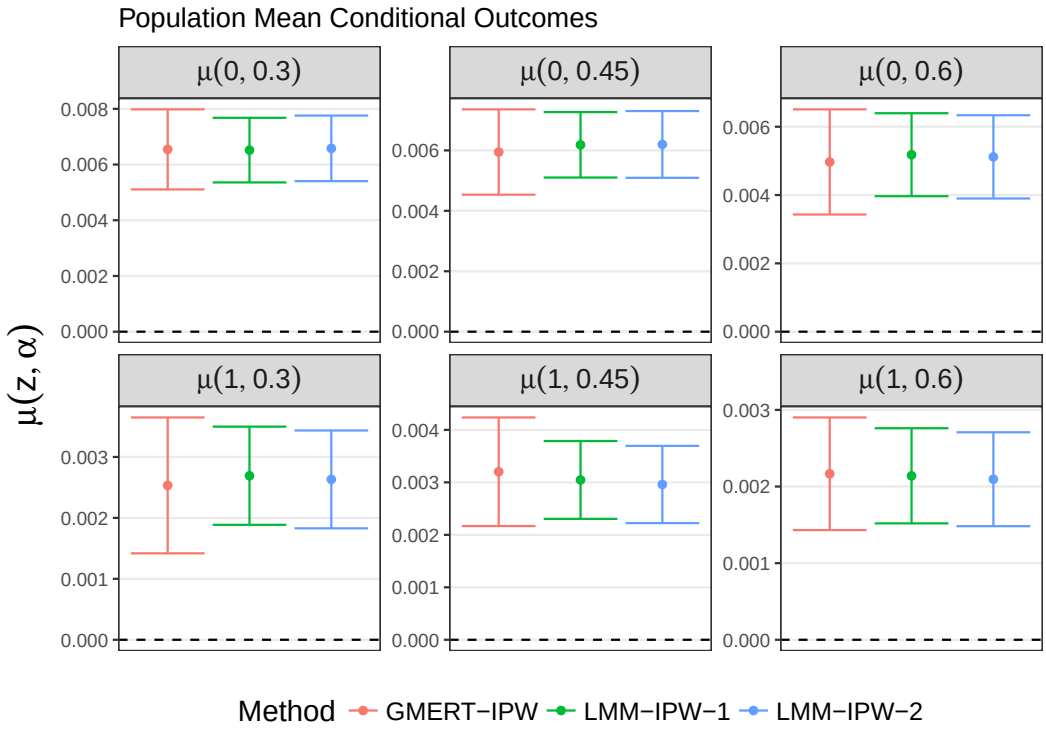
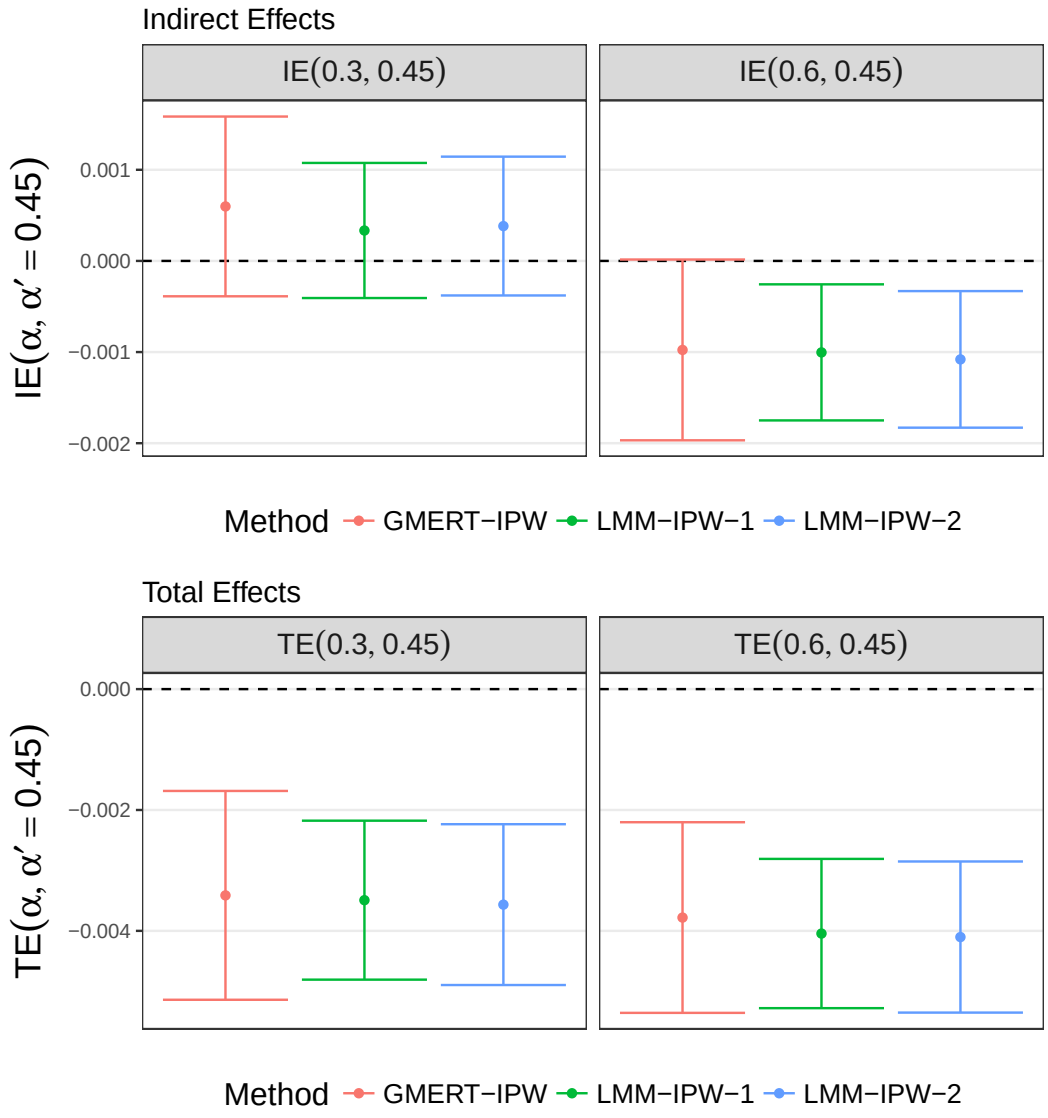


Figure A.6: Estimates and 95% confidence intervals for Indirect and Total Effects.



A.2 Sketch of the GMERT algorithm

A brief sketch of the GMERT algorithm is provided here. The algorithm relies on the PQL method to create a continuously-valued working response, which is then "de-correlated" by subtracting empirical BLUPs. Let $\mathcal{S}$ be a sample of data of $i = 1, \ldots, M$ i.i.d. clusters, where cluster i has $j = 1, \ldots, N_i$ individuals. Let $g(x) = \mathcal{L}(x)$ be the logit-link function for binary response such that $\mu_i = \mathbb{E}(A_i|L_i) = g^{-1}(\eta_i)$.

Step (0). Initialization.

- Set $\Theta_{(0)}(L) = L$ to be the identity function. Thus there are fixed effects $\beta_{\Theta_{(0)}}$ and random component σ such that $b \sim N(0, \sigma_{\Theta_{(0)}})$.
- Estimate eBLUPs $\hat{b}_i^{(0)}$

Step (s). Until convergence, do:

- Estimate the linear predictor, $\hat{\eta}_i^{(s)} = \Theta_{(s-1)}(L_i)^\intercal \hat{\beta}_{\Theta_{(s-1)}} + \hat{b}_i^{(s-1)}$
- Calculate the conditional mean response, $\hat{\mu}_i^{(s)} = g^{-1}(\hat{\eta}_i^{(s)})$
- "Linearize" the response by computing the "working response" with a Taylor Series approximation, $\tilde{A}_i^{(s)} := g(\hat{\mu}_i^{(s)}) + (A_i - \hat{\mu}_i^{(s)}) g'(\hat{\mu}_i^{(s)})$
- Compute the working weights, $w_{ij}^{(s)} = \left[v_{ij} \left\{ g'(\hat{\mu}_{ij}^{(s)}) \right\}^2 \right]^{-1}$, where $v_{ij} = v(\mu_{ij})$ and $v(\cdot)$ is the known variance function for the exponential family.
- Calculate the "adjusted working response" by subtracting the previous eBLUPs to "de-correlate" the response, $A_i^{(s)} = \tilde{A}_i^{(s-1)} - \hat{b}_i^{(s-1)}$
- Fit a CART regression tree algorithm (weighted by $w_{ij}^{(s)}$) to predict the adjusted working response $A_{ij}^{(s)}$ for predictors L_{ij} for every individual j in every group i. Let $\Theta_{(s)}$ be the partition that is recovered from the tree algorithm, and define $\hat{\beta}_{\Theta_{(s)}}$ to be the predicted probabilities for the terminal nodes.
- Calculate eBLUPs $\hat{b}_i^{(s)}$ from partition $\Theta_{(s)}$ based on above, as well as $\hat{\sigma}^{(s)}$ and the variance component for the fixed effects.
- If convergence criteria is met, then define $\Theta = \Theta_{(s)}$ to be the partition, and exit the loop.

APPENDIX B: TECHNICAL DETAILS FOR CHAPTER 3

B.1 Asymptotic Properties

Provided here are details for the consistency and asymptotic normality of the proposed estimators. In subsection B.1.1 we derive the estimator's decomposed form, $\hat{\theta}_M(\alpha) = \overline{\theta(\alpha|X)} + E_M(\alpha) + B_M(\alpha)$, as in (3.31). These components are key to establishing consistency in subsection B.1.2 and asymptotic normality in subsection B.1.3 under the assumptions from Theorem 3.5.1.

B.1.1 Decomposition

First, we reformulate the estimator from the form provided in (3.29). Write

$$\hat{\theta}_M(\alpha) = N^{-1}\sum_{i=1}^{N}\pi(A_i,\alpha)\bar{Y}_i + N^{-1}\sum_{i=1}^{N}\sum_{a\in\mathcal{A}(c)} M^{-1}\left[\sum_{j\in J_M(i,a)}\bar{Y}_j I(A_i\neq a)\right]\pi(a,\alpha)$$

The second term on the right side of above can be written as

$$N^{-1}\sum_{a\in\mathcal{A}(c)}\pi(a,\alpha)M^{-1}\sum_{i=1}^{N}\left[\sum_{j\in J_M(i,a)}\bar{Y}_j I(A_i\neq a)\right]$$

Considering just the last two summands of this term,

$$\begin{aligned}\sum_{i=1}^{N}\left[\sum_{j\in J_M(i,a)}\bar{Y}_j I(A_i\neq a)\right] &= \sum_{i=1}^{N}\sum_{j=1}^{N} I(j\in J_M(i,a))\bar{Y}_j I(A_j=a)\\ &= \sum_{j=1}^{N}\bar{Y}_j I(A_j=a)\sum_{i=1}^{N} I(j\in J_M(i,a))\end{aligned}$$

which by definition equals $\sum_{j=1}^{N} \bar{Y}_j I(A_j = a) K_M(j)$. Now,

$$\hat{\theta}_M(\alpha) = N^{-1} \sum_{i=1}^{N} \sum_{a \in \mathcal{A}(c)} \pi(a, \alpha) \left[\bar{Y}_i I(A_i = a) + \bar{Y}_i I(A_i = a) \frac{K_M(i)}{M} \right],$$

and so:

$$\hat{\theta}_M(\alpha) = \frac{1}{N} \sum_{i=1}^{N} \Big(1 + \frac{K_M(i)}{M} \Big) \bar{Y}_i \pi(A_i, \alpha). \tag{B.1}$$

For the Overall Effect estimator, since $\hat{\theta}_M(\alpha, \alpha') = \hat{\theta}_M(\alpha) - \hat{\theta}_M(\alpha')$, then

$$\hat{\theta}_M(\alpha, \alpha') = N^{-1} \sum_{i=1}^{N} \sum_{a \in \mathcal{A}(c)} \left(1 + \frac{K_M(i)}{M} \right) \bar{Y}_i \left\{ \pi(A_i, \alpha) - \pi(A_i, \alpha') \right\}.$$

The three components on the right side of (3.31) can be written in terms that come from each unit in the study sample. For example, each unit's contribution to the weighted residual term E_M can be written as $E_M(\alpha) = \frac{1}{N} \sum_{i=1}^{N} E_{M,i}(\alpha)$ where

$$E_{M,i}(\alpha) = \Big(1 + \frac{K_M(i)}{M} \Big) \epsilon_i \pi(A_i, \alpha). \tag{B.2}$$

Similarly, the sample average conditional mean outcome can be written as the mean $\overline{\theta(\alpha|X)} = \frac{1}{N} \sum_{i=1}^{N} \theta(\alpha|X_i)$ where

$$\theta(\alpha|X_i) = \sum_{a \in \mathcal{A}(c)} \mu_a(X_i) \pi(a, \alpha). \tag{B.3}$$

Define the matching discrepancy for the m^{th} match for unit i to be $U_{i,m,a} = X_{j_m(i,a)} - X_i$, and define the bias arising from the matching discrepancy to be $B_{i,m,a} = \mu_a(X_{j_m(i,a)}) - \mu_a(X_i)$. Then we can rewrite the conditional bias term in (3.34) as:

$$B_M(\alpha) = \frac{1}{N} \frac{1}{M} \sum_{i=1}^{N} \sum_{m=1}^{M} \sum_{a \in \mathcal{A}(c)} B_{i,m,a} \pi(a, \alpha) \tag{B.4}$$

We show $\hat{\theta}_M(\alpha) - E_M(\alpha) - \overline{\theta(\alpha|X)} = B_M(\alpha)$. By adding and subtracting copies of $\mu_{A_i}(X_i)\pi(A_i,\alpha)$ and $\sum_{a\neq A_i}\mu_a(X_i)\pi(a,\alpha)$ from $\bar{Y}_i\pi(A_i,\alpha)$ in (B.1), then $\hat{\theta}_M(\alpha)$ equals

$$\frac{1}{N}\sum_{i=1}^{N}\Big(1+\frac{K_M(i)}{M}\Big)\Big[\epsilon_i\pi(A_i,\alpha)+\sum_{a\in\mathcal{A}(c)}\mu_a(X_i)\pi(a,\alpha)-\sum_{a\neq A_i}\mu_a(X_i)\pi(a,\alpha)\Big],$$

and so $\hat{\theta}_M(\alpha) - E_M(\alpha) - \overline{\theta(\alpha|X)}$ equals

$$\frac{1}{N}\sum_{i=1}^{N}\Big(\frac{K_M(i)}{M}\Big)\Big[\sum_{a\in\mathcal{A}(c)}\mu_a(X_i)\pi(a,\alpha)\Big]-\frac{1}{N}\sum_{i=1}^{N}\Big(1+\frac{K_M(i)}{M}\Big)\Big[\sum_{a\neq A_i}\mu_a(X_i)\pi(a,\alpha)\Big].$$

Consider just the second term:

$$\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{N}\Big(1+\frac{K_M(i)}{M}\Big)\Big[\sum_{a\neq A_i}\mu_a(X_i)\pi(a,\alpha)\Big]\\
&=\frac{1}{N}\sum_{i=1}^{N}\Big[\sum_{a\neq A_i}\mu_a(X_i)\pi(a,\alpha)\Big]+\frac{1}{N}\sum_{i=1}^{N}\Big(\frac{K_M(i)}{M}\Big)\Big[\sum_{a\neq A_i}\mu_a(X_i)\pi(a,\alpha)\Big]\\
&=\frac{1}{N}\sum_{i=1}^{N}\Big(\frac{K_M(i)}{M}\Big)\Big[\sum_{a\neq A_i}\mu_a(X_i)\pi(a,\alpha)\Big]+\frac{1}{N}\sum_{i=1}^{N}\Big[\sum_{a\neq A_i}\mu_a(X_i)\pi(a,\alpha)\Big].
\end{aligned}$$

Then $\hat{\theta}_M(\alpha) - E_M(\alpha) - \overline{\theta(\alpha|X)}$ becomes

$$\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{N}\Big(\frac{K_M(i)}{M}\Big)\Big[\sum_{a\in\mathcal{A}(c)}\mu_a(X_i)\pi(a,\alpha)-\sum_{a\neq A_i}\mu_a(X_i)\pi(a,\alpha)\Big]-\\
&\qquad\frac{1}{N}\sum_{i=1}^{N}\Big[\sum_{a\neq A_i}\mu_a(X_i)\pi(a,\alpha)\Big]\\
&=\frac{1}{N}\sum_{i=1}^{N}\Big(\frac{K_M(i)}{M}\Big)\mu_{A_i}(X_i)\pi(A_i,\alpha)-\frac{1}{N}\sum_{i=1}^{N}\sum_{a\neq A_i}\mu_a(X_i)\pi(a,\alpha).
\end{aligned}$$

Expand the first term:

$$\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{N}\Big(\frac{K_M(i)}{M}\Big)\mu_{A_i}(X_i)\pi(A_i,\alpha) \\
&= \frac{1}{N}\sum_{i=1}^{N}\frac{1}{M}\sum_{l=1}^{N}\sum_{m=1}^{M}\sum_{a\in\mathcal{A}(c)} I(i=j_m(l,a))\mu_a(X_{j_m(l,a)})\pi(a,\alpha)I(A_i=a) \\
&= \frac{1}{N}\frac{1}{M}\sum_{l=1}^{N}\sum_{m=1}^{M}\sum_{a\in\mathcal{A}(c)} \mu_a(X_{j_m(l,a)})\pi(a,\alpha)I(A_l\neq a)\sum_{i=1}^{N} I(i=j_m(l,a))I(A_i=a).
\end{aligned}$$

Now consider the term $\sum_{i=1}^{N} I(i=j_m(l,a))I(A_i=a)$. For a fixed $l\leq N, a\in\mathcal{A}(c)$, and $m\leq M$, then this sum equals exactly 1. Then

$$\frac{1}{N}\sum_{i=1}^{N}\Big(\frac{K_M(i)}{M}\Big)\mu_{A_i}(X_i)\pi(A_i,\alpha) = \frac{1}{N}\frac{1}{M}\sum_{i=1}^{N}\sum_{m=1}^{M}\sum_{a\in\mathcal{A}(c)} \mu_a(X_{j_m(i,a)})\pi(a,\alpha)I(A_i\neq a).$$

Finally we reach (3.31) as $\hat{\theta}_M(\alpha) - E_M(\alpha) - \overline{\theta(\alpha|X)}$ equals

$$\begin{aligned}
&\frac{1}{N}\frac{1}{M}\sum_{i=1}^{N}\sum_{m=1}^{M}\sum_{a\in\mathcal{A}(c)} \mu_a(X_{j_m(i,a)})\pi(a,\alpha)I(A_i\neq a) - \frac{1}{N}\sum_{i=1}^{N}\sum_{a\neq A_i}\mu_a(X_i)\pi(a,\alpha) \\
&= \frac{1}{N}\frac{1}{M}\sum_{i=1}^{N}\sum_{m=1}^{M}\sum_{a\in\mathcal{A}(c)} \big[\mu_a(X_{j_m(i,a)}) - \mu_a(X_i)\big]\pi(a,\alpha)I(A_i\neq a) \;=\; B_M(\alpha).
\end{aligned}$$

B.1.2 Consistency

We first present a proof regarding the consistency of the conditional bias term.

Theorem B.1.1. ***Asymptotic Bounds of Conditional Bias:*** *Assume the first three assumptions as in Theorem 3.5.1, and the smoothness assumption that $\mu(x,a)$ and $\sigma^2(x,a)$ are Lipschitz in $\mathbb{X}$. Then $B_M(\alpha) = O_p(N^{-1/k})$.*

Proof. We show the second moment of $N^{1/k}B_M(\alpha)$ is bounded in probability. First,

$$\begin{aligned}
&\mathbb{E}[(N^{1/k}B_M(\alpha))^2] \\
&= N^{2/k}\,\mathbb{E}\left[\left(\frac{1}{N}\frac{1}{M}\sum_{i=1}^{N}\sum_{m=1}^{M}\sum_{a\in\mathcal{A}(c)} B_{i,m,a}\pi(a,\alpha)I(A_i \neq a)\right)^2\right] \\
&= N^{2/k}\,\mathbb{E}\left[\frac{1}{N^2}\frac{1}{M^2}\left(\sum_{i=1}^{N}\sum_{m=1}^{M}\sum_{a\in\mathcal{A}(c)} B_{i,m,a}\pi(a,\alpha)I(A_i \neq a)\right)^2\right] \\
&\leq N^{2/k}\,\mathbb{E}\left[|\mathcal{A}(c)|^2 \arg\max_{i,m,a} |B_{i,m,a}|^2\right]
\end{aligned}$$

So then $\mathbb{E}[(N^{1/k}B_M(\alpha))^2] \leq N^{2/k}|\mathcal{A}(c)|^2 L_1^* \,\mathbb{E}[\arg\max ||U_{i,m,a}||^2]$ where L_1^* is some constant for the Lipschitz inequality. Lemma 1 from Abadie and Imbens (2006) states that $\left(N^{1/k}||U_{m,i,a}||\right)$ has bounded moments, so $\mathbb{E}[N^{2/k}B_M(\alpha)^2] = O_p(1)$. Applying Markov's inequality finishes the proof. □

Theorem B.1.2. ***Consistency of the Proposed Estimator****: Assume the first three assumptions as in Theorem 3.5.1, and the smoothness assumption that $\mu(x,a)$ and $\sigma^2(x,a)$ are Lipschitz in $\mathbb{X}$. Then $\hat{\theta}_M(\alpha) - \theta(\alpha) \xrightarrow{p} 0$*

Proof. First consider $\overline{\theta(\alpha|X)} = \frac{1}{N}\sum_{i=1}^{N}\theta(\alpha|X_i)$. The random variable $(\theta(\alpha|X) - \theta(\alpha))$ has mean zero and finite variance. By assumption, $\mu_a(x)$ is bounded over $x \in \mathbb{X}$ and $a \in \mathcal{A}(c)$, so then $\overline{\theta(\alpha|X)} - \theta \xrightarrow{p} 0$ by the LLN.

For the residual term, since $\mathbb{E}(\epsilon_i \epsilon_j | X, A) = 0$ for $i \neq j$, then:

$$
\begin{aligned}
& \mathbb{E}\left[\left\{N^{1/2} E_M(\alpha)\right\}^2\right] \\
&= \frac{1}{N} \mathbb{E}\left[\sum_{i=1}^{N}\sum_{j=1}^{N} \pi(A_i,\alpha)\pi(A_j,\alpha)\left(1+\frac{K_M(i)}{M}\right)\left(1+\frac{K_M(i)}{M}\right)\epsilon_i\epsilon_j\right] \\
&= \frac{1}{N} \mathbb{E}\left[\sum_{i=1}^{N}\sum_{j=1}^{N} \pi(A_i,\alpha)\pi(A_j,\alpha)\left(1+\frac{K_M(i)}{M}\right)\left(1+\frac{K_M(j)}{M}\right)\mathbb{E}\left(\epsilon_i\epsilon_j | X, A\right)\right] \\
&= \frac{1}{N} \sum_{i=1}^{N} \mathbb{E}\left[\pi(A_j,\alpha)^2\left(1+\frac{K_M(i)}{M}\right)^2 \sigma^2(X_i, A_i)\right] \\
&= \mathbb{E}\left[\pi(A_j,\alpha)^2\left(1+\frac{K_M(i)}{M}\right)^2 \sigma^2(X_i, A_i)\right].
\end{aligned}
$$

Lemma 3 of Abadie and Imbens (2006) states that the moments of $K_M(i)$ are bounded uniformly in N, and so $\mathbb{E}[\{N^{1/2} E_M(\alpha)\}^2] = O_p(1)$. Applying Markov's inequality shows that $E_M(\alpha) = o_p(1)$. Applying Theorem B.1.1 completes the proof. □

B.1.3 Asymptotic Normality

Ignoring the conditional bias, we derive the asymptotic variance by working with each of the two terms in

$$
\mathrm{Var}(\hat{\theta}_M(\alpha)) = \mathbb{E}\left[\mathrm{Var}\left(\hat{\theta}_M(\alpha)|X, A\right)\right] + \mathrm{Var}\left[\mathbb{E}\left(\hat{\theta}_M(\alpha)|X, A\right)\right].
$$

We apply (B.3) to show that $\mathrm{Var}\left[\mathbb{E}\left(\hat{\theta}_M(\alpha)|X, A\right)\right]$ equals

$$
\begin{aligned}
& \mathrm{Var}\left[\overline{\theta(\alpha|X)}\right] \\
&= \frac{1}{N^2}\sum_{i=1}^{N} \mathrm{Var}\left[\theta(\alpha|X_i)\right] \; + \; \frac{1}{N^2}\sum_{i\neq j} Cov(\theta(\alpha|X_i), \theta(\alpha|X_j)) \\
&= \frac{1}{N} \mathrm{Var}\left[\theta(\alpha|X_i)\right] \; + \; 0.
\end{aligned}
$$

So now $N \operatorname{Var}\left[\mathbb{E}\left(\hat{\theta}_M(\alpha)|X,A\right)\right] = V^{\theta(\alpha|X)}$ from (3.35). For $\mathbb{E}\left[\operatorname{Var}\left(\hat{\theta}_M(\alpha)|X,A\right)\right]$ consider the conditional variance inside the expectation. Using the reformulated estimator provided in (B.1), $\operatorname{Var}\left(\hat{\theta}_M(\alpha)|X,A\right)$ equals

$$
\begin{aligned}
&\operatorname{Var}\left(\frac{1}{N}\sum_{i=1}^{N}\left(1+\frac{K_M(i)}{M}\right)\bar{Y}_i\pi(A_i,\alpha)|X,A\right) \\
&= \frac{1}{N^2}\sum_{i=1}^{N}\left(1+\frac{K_M(i)}{M}\right)^2\pi(A_i,\alpha)^2\operatorname{Var}\left(\bar{Y}_i|X,A\right) \;+ \\
&\quad \frac{1}{N^2}\sum_{i\neq j}\left(1+\frac{K_M(i)}{M}\right)\left(1+\frac{K_M(j)}{M}\right)\pi(A_i,\alpha)\pi(A_j,\alpha)Cov\left(\bar{Y}_i,\bar{Y}_j|X,A\right) \\
&= \frac{1}{N^2}\sum_{i=1}^{N}\left(1+\frac{K_M(i)}{M}\right)^2\pi(A_i,\alpha)^2\sigma^2(X_i,A_i) \;+\; 0
\end{aligned}
$$

Now we have that $N \operatorname{Var}\left(\hat{\theta}_M(\alpha)|X,A\right) = V^{E_M(\alpha)}$ from (3.37). Combining the previous results, we arrive at

$$\operatorname{Var}(\hat{\theta}_M(\alpha)) = \frac{1}{N}\left[\mathbb{E}\,V^{E_M(\alpha)} + V^{\theta(\alpha|X)}\right].$$

Abadie and Imbens (2006) sometimes refer to this term as the "marginal" variance.

We consider the distribution of the $E_{M,i}(\alpha)$ terms from (B.2) conditional on treatment A and pre-treatment covariates X. Clearly $\mathbb{E}(E_{M,i}(\alpha)|X,A) = 0$ because all are constants except for $\mathbb{E}(\epsilon_i|X,A) = 0$. We also note that $E_{M,i}(\alpha) \perp E_{M,j}(\alpha)|\{X,A\}$ because the residuals are conditionally independent. The variances of these terms are nonidentical because

$$\operatorname{Var}(E_{M,i}(\alpha)|X,A) = \left(1+\frac{K_M(i)}{M}\right)^2\pi(A_i,\alpha)^2\sigma^2(X_i,A_i)$$

depends on $\sigma^2(X_i,A_i)$. We use the Lindeberg-Feller theorem to show that these

independent-non-identically-distributed terms converge to a standard normal distribution (conditional on X and A).

Theorem B.1.3. ***Lindeberg-Feller CLT for Weighted Residual Term:*** *Under the assumptions of Theorem 3.5.1, the following term vanishes:*

$$\frac{1}{NV^{E_M(\alpha)}} \sum_{i=1}^{N} \mathbb{E}\left[(E_{M,i}(\alpha))^2 I(|E_{M,i}(\alpha)| > \eta\sqrt{NV^{E_M(\alpha)}}) \middle| X, A\right]. \tag{B.5}$$

Then, the Lindeberg condition is satisfied, and $(N/V^{E,\alpha})^{1/2} E_M(\alpha) \xrightarrow{d} N(0,1)$.

Proof. Considering only the summand,

$$\begin{aligned}
&\mathbb{E}\left[(E_{M,i}(\alpha))^2 I(|E_{M,i}(\alpha)| > \eta\sqrt{NV^{E_M(\alpha)}}) \middle| X, A\right] \\
&\leq \left(\mathbb{E}\left[(E_{M,i}(\alpha))^4 \middle| X, A\right]\right)^{1/2} \left(\mathbb{E}\left[I(|E_{M,i}(\alpha)| > \eta\sqrt{NV^{E_M(\alpha)}})^1 \middle| X, A\right]\right)^{1/2} \\
&= \left(\mathbb{E}\left[(E_{M,i}(\alpha))^4 \middle| X, A\right]\right)^{1/2} \left(\Pr\left[|E_{M,i}(\alpha)| > \eta\sqrt{NV^{E_M(\alpha)}}) \middle| X, A\right]\right)^{1/2} \\
&\leq \left(\mathbb{E}\left[(E_{M,i}(\alpha))^4 \middle| X, A\right]\right)^{1/2} \left(\frac{\mathbb{E}\left[(E_{M,i}(\alpha))^2 | X, A\right]}{(\eta\sqrt{NV^{E_M(\alpha)}})^2}\right)^{1/2} \\
&= \frac{1}{\eta\sqrt{NV^{E_M(\alpha)}}} \left(\mathbb{E}\left[(E_{M,i}(\alpha))^4 \middle| X, A\right]\right)^{1/2} \left(\mathbb{E}\left[(E_{M,i}(\alpha))^2 | X, A\right]\right)^{1/2}
\end{aligned}$$

Let $\underline{\sigma}^2(\alpha) = \inf_{X,A} \sigma^2(X,A)\pi(A,\alpha)$; then $V^{E_M(\alpha)} \geq \underline{\sigma}^2(\alpha)$. Also define $\overline{\sigma}^2 = \sup_{X,A} \sigma^2(X,A)$. Finally define $\overline{C}^2 = \sup_{X,A} \mathbb{E}(\epsilon_i^4 | X, A) < \infty$. We now have that

$$\begin{aligned}
&\frac{1}{NV^{E_M(\alpha)}} \sum_{i=1}^{N} \mathbb{E}\left[(E_{M,i}(\alpha))^2 I(|E_{M,i}(\alpha)| > \eta\sqrt{NV^{E_M(\alpha)}}) \middle| X, A\right] \\
&\leq \frac{1}{\sqrt{N}} \frac{1}{\eta \left(V^{E_M(\alpha)}\right)^{3/2}} \frac{1}{N} \sum_{i=1}^{N} \left(\mathbb{E}\left[(E_{M,i}(\alpha))^4 \middle| X, A\right]\right)^{1/2} \left(\mathbb{E}\left[(E_{M,i}(\alpha))^2 | X, A\right]\right)^{1/2} \\
&\leq \frac{1}{\sqrt{N}} \frac{1}{\eta\underline{\sigma}^3(\alpha)} \frac{1}{N} \sum_{i=1}^{N} \left(1 + \frac{K_M(i)}{M}\right)^3 \pi(A_i,\alpha)^3 \left(\mathbb{E}\left[\epsilon_i^4 \middle| X, A\right]\right)^{1/2} \left(\mathbb{E}\left[\epsilon_i^2 | X, A\right]\right)^{1/2}.
\end{aligned}$$

Now, (B.5) is bounded above by

$$\frac{1}{\sqrt{N}}\frac{\overline{\sigma}\overline{C}}{\eta\underline{\sigma}^3(\alpha)}\frac{1}{N}\sum_{i=1}^{N}\left(1+\frac{K_M(i)}{M}\right)^3.$$

Since the moments of $K_M(i)$ are bounded uniformly in N by Lemma 3 of Abadie and Imbens (2006), then $\mathbb{E}[(1+K_M(i)/M)^3] = O_p(1)$. From the LLN,

$$\frac{1}{N}\sum_{i=1}^{N}\left(1+\frac{K_M(i)}{M}\right)^3 \to \mathbb{E}\left[\left(1+\frac{K_M(i)}{M}\right)^3\right] = O_p(1).$$

Since $\overline{\sigma}\overline{C}/\{\eta\underline{\sigma}^3(\alpha)\}$ is a constant, then (B.5) vanishes:

$$\frac{1}{NV^{E_M(\alpha)}}\sum_{i=1}^{N}\mathbb{E}\left[(E_{M,i}(\alpha))^2 I(|E_{M,i}(\alpha)| > \eta\sqrt{NV^{E_M(\alpha)}})\big|X,A\right] \leq \frac{1}{\sqrt{N}}O_p(1) = o_p(1),$$

the Lindeberg condition is satisfied, and

$$\frac{1}{\sqrt{NV^{E_M(\alpha)}}}\sum_{i=1}^{N}E_{M,i}(\alpha) = \frac{\sqrt{N}}{\sqrt{V^{E,\alpha}}}E_M(\alpha) \xrightarrow{d} N(0,1).$$

□

B.2 Estimating Asymptotic Variance

In this supplement we derive an estimator for the variance of the conditional mean, $V^{\theta(\alpha|X)}$. This estimator, like the estimator of $V^{E_M(\alpha)}$, relies on the estimator $\hat{\sigma}^2_J(x,a)$ of $\sigma^2(x,a)$ from Abadie and Imbens (2006), which we briefly describe in Section B.2.1. In Section B.2.2 we derive an approximation for $V^{\theta(\alpha|X)}$; this approximation is used in Section B.2.3 to introduce the estimator $\hat{V}^{\theta(\alpha|X)}_{M,J}$ as shown in the main paper.

B.2.1 Estimating the Conditional Outcome Variance

We use the within-treatment-level matching estimator proposed in Abadie and Imbens (2006) to estimate $\sigma^2(X, A)$, which they term the "conditional outcome variance." Define $l_m(i)$ to be the mth closest unit to unit i among the units with the same treatment, A_i. For some fixed (investigator-supplied) $J \geq 1$, define

$$\hat{\sigma}^2_J(X_i, A_i) = \frac{J}{J+1}\left(\bar{Y}_i - \frac{1}{J}\sum_{m=1}^{J} \bar{Y}_{l_m(i)}\right)^2. \tag{B.6}$$

Abadie and Imbens (2006) remark that in practice this is not necessarily consistent for $\sigma^2(X_i, A_i)$. Theorems 6 and 7 from Abadie and Imbens (2006) state that this within-treatment-level matching estimator results in consistent estimates of the variance of the estimator of the Average Treatment Effect under certain scenarios.

B.2.2 Approximation of the Variance of the Conditional Mean

$$\mathbb{E}\left[\left(\sum_{a\in\mathcal{A}(c)} \tilde{Y}_i(a)\pi(a,\alpha) - \theta(\alpha)\right)^2\right] \approx \mathbb{E}\left[\epsilon_i^2\pi(A_i,\alpha)^2 + \frac{1}{M^2}\sum_{m=1}^{M}\sum_{a\neq A_i}\epsilon^2_{j_m(i,a)}\pi(a,\alpha)^2\right] + V^{\theta(\alpha|X_i)}, \tag{B.7}$$

where $V^{\theta(\alpha|X_i)}$ is given in (3.35).

The term on the left side of (B.7) equals:

$$\mathbb{E}\left[\left(\sum_{a\in\mathcal{A}(c)}\mu_a(X_i)\pi(a,\alpha)-\theta(\alpha)+\sum_{a\in\mathcal{A}(c)}(\tilde{Y}_i(a)-\mu_a(X_i))\pi(a,\alpha)\right)^2\right]$$
$$=\mathbb{E}\left[\left(\sum_{a\in\mathcal{A}(c)}\mu_a(X_i)\pi(a,\alpha)-\theta(\alpha)\right)^2\right]+\mathbb{E}\left[\left(\sum_{a\in\mathcal{A}(c)}(\tilde{Y}_i(a)-\mu_a(X_i))\pi(a,\alpha)\right)^2\right]+$$
$$\mathbb{E}\left[\left(\sum_{a\in\mathcal{A}(c)}\mu_a(X_i)\pi(a,\alpha)-\theta(\alpha)\right)\left(\sum_{a\in\mathcal{A}(c)}(\tilde{Y}_i(a)-\mu_a(X_i))\pi(a,\alpha)\right)\right]. \quad \text{(B.8)}$$

By (3.35), the first term in (B.8) is simply $\mathbb{E}[\{\sum\mu_a(X_i)\pi(a,\alpha)-\theta(\alpha)\}^2]=V^{\theta(\alpha|X)}$. The second term in (B.8), $\mathbb{E}[\{\sum(\tilde{Y}_i(a)-\mu_a(X_i))\pi(a,\alpha)\}^2]$, equals

$$\mathbb{E}\left[\left((\bar{Y}_i-\mu_{A_i}(X_i))\pi(A_i,\alpha)+\sum_{a\neq A_i}(\tilde{Y}_i(a)-\mu_a(X_i))\pi(a,\alpha)\right)^2\right]$$
$$=\mathbb{E}\left[\left(\epsilon_i\pi(A_i,\alpha)+\sum_{a\neq A_i}\frac{1}{M}\sum_{m=1}^{M}\epsilon_{j_m(i,a)}\pi(a,\alpha)+\right.\right.$$
$$\left.\left.\sum_{a\neq A_i}\frac{1}{M}\sum_{m=1}^{M}\left\{\mu_a(X_{j_m(i,a)})-\mu_a(X_i)\right\}\pi(a,\alpha)\right)^2\right]. \quad \text{(B.9)}$$

Group the first two terms from the right side of (B.9) together, then expand the square inside the expectation. Consider the third term resulting from the expanded (B.9),

$$\mathbb{E}\left[\left(\sum_{a\neq A_i}\frac{1}{M}\sum_{m=1}^{M}(\mu_a(X_{j_m(i,a)})-\mu_a(X_i))\pi(a,\alpha)\right)^2\right].$$

When N is large and matching discrepancy is small such that $||X_i-X_{j_m(i,a)}||\approx 0$, the Lipschitz property of the regression functions μ_a implies that $|\mu_a(X_{j_m(i,a)})-\mu_a(X_i)|\approx 0$, and so the above term is then approximately zero.

For the first term resulting from the expanded (B.9), note that

$$\begin{aligned}
&\mathbb{E}\left[\left(\epsilon_i\pi(A_i,\alpha)+\sum_{a\neq A_i}\frac{1}{M}\sum_{m=1}^{M}\epsilon_{j_m(i,a)}\pi(a,\alpha)\right)^2\right] \\
&\quad=\mathbb{E}\left[\epsilon_i^2\pi(A_i,\alpha)^2\right]+\mathbb{E}\left[\left(\sum_{a\neq A_i}\frac{1}{M}\sum_{m=1}^{M}\epsilon_{j_m(i,a)}\pi(a,\alpha)\right)^2\right]- \\
&\qquad 2\,\mathbb{E}\left[\sum_{a\neq A_i}\frac{1}{M}\sum_{m=1}^{M}\epsilon_i\epsilon_{j_m(i,a)}\pi(A_i,\alpha)\pi(a,\alpha)\right] \qquad \text{(B.10)}
\end{aligned}$$

Now, for the second term on the right side of (B.10), expand the square and iterate expectations on X, W. Since $\mathbb{E}\left(\epsilon_i\epsilon_j|X,W\right)=0$ when $i\neq j$, then:

$$\mathbb{E}\left[\left(\sum_{a\neq A_i}\frac{1}{M}\sum_{m=1}^{M}\epsilon_{j_m(i,a)}\pi(a,\alpha)\right)^2\right]=\mathbb{E}\left[\frac{1}{M^2}\sum_{a\neq A_i}\sum_{m=1}^{M}\epsilon_{j_m(i,a)}^2\pi(a,\alpha)^2\right].$$

The third term of the right side in (B.10) equals zero as $\mathbb{E}\left(\epsilon_i\epsilon_{j_M(i,a)}|X,W\right)=0$. Lastly, the cross product of the two grouped terms from the expanded (B.9) can be shown to be zero by iterated expectations with respect to X, W. Now all of (B.9) is approximately equal to the first term of the right side of (B.7).

Finally, we make use of a similar approximation to show that the third and final term from the right side of (B.8) is approximately zero. Iterated expectations of

$$\mathbb{E}\left[\left(\sum_{a\in\mathcal{A}(c)}\mu_a(X_i)\pi(a,\alpha)-\theta(\alpha)\right)\left(\sum_{a\in\mathcal{A}(c)}\left(\tilde{Y}_i(a)-\mu_a(X_i)\right)\pi(a,\alpha)\right)\right],$$

shows that the second term in the product equals

$$\mathbb{E}\left(\left(\bar{Y}_i-\mu_{A_i}(X_i)\right)\pi(a,\alpha)+\sum_{a\neq A_i}\left(\tilde{Y}_i(a)-\mu_a(X_i)\right)\pi(a,\alpha)\middle|X,A\right).$$

Note that $\mathbb{E}(\bar{Y}_i - \mu_{A_i}(X_i)) = 0$ and

$$\mathbb{E}(\tilde{Y}_i(a)|X, A) = \mathbb{E}\left(\frac{1}{M}\sum_{m=1}^{M}\bar{Y}_{j_m(i,a)}\Big|X, A\right) = \frac{1}{M}\sum_{m=1}^{M}\mu_a(X_{j_m(i,a)}),$$

and so the third term on the right side of (B.8) is:

$$\mathbb{E}\left[\left(\sum_{a\in\mathcal{A}(c)}\mu_a(X_i)\pi(a,\alpha) - \theta(\alpha)\right)\left(\sum_{a\neq A_i}\left(\frac{1}{M}\sum_{m=1}^{M}\mu_a(X_{j_m(i,a)}) - \mu_a(X_i)\right)\pi(a,\alpha)\right)\right].$$

Again, since $|\mu_a(X_{j_m(i,a)}) - \mu_a(X_i)| \approx 0$, then this is approximately zero. Combining all the above results in the desired approximation of $V^{\theta(\alpha|X)}$ in (B.7).

B.2.3 Estimator of the Variance of the Conditional Mean

For the second term in (B.7), note that

$$\begin{aligned}
&\mathbb{E}\left[\epsilon_i^2\pi(A_i,\alpha)^2 + \frac{1}{M^2}\sum_{m=1}^{M}\sum_{a\neq A_i}\epsilon_{j_m(i,a)}^2\pi(a,\alpha)^2\right] \\
&= \mathbb{E}\left[\mathbb{E}\left(\epsilon_i^2\pi(A_i,\alpha)^2 + \frac{1}{M^2}\sum_{m=1}^{M}\sum_{a\neq A_i}\epsilon_{j_m(i,a)}^2\pi(a,\alpha)^2\Big|X, W\right)\right] \\
&= \mathbb{E}\left[\mathbb{E}\left(\epsilon_i^2\Big|X, W\right)\pi(A_i,\alpha)^2 + \frac{1}{M^2}\sum_{m=1}^{M}\sum_{a\neq A_i}\mathbb{E}\left(\epsilon_{j_m(i,a)}^2\Big|X, W\right)\pi(a,\alpha)^2\right] \\
&= \mathbb{E}\left[\sigma^2(X_i, A_i)\pi(A_i,\alpha)^2 + \frac{1}{M^2}\sum_{m=1}^{M}\sum_{a\neq A_i}\sigma^2(X_{j_m(i,a)}, A_{j_m(i,a)})\pi(a,\alpha)^2\right]
\end{aligned}$$

We can estimate the $\sigma^2(x,w)$ components by $\hat{\sigma}^2_J(x,w)$, as described in Section B.2.1. So, we estimate the above with

$$
\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{N}\left[\hat{\sigma}^2_J(X_i,A_i)\pi(A_i,\alpha)^2+\frac{1}{M^2}\sum_{m=1}^{M}\sum_{a\neq A_i}\hat{\sigma}^2_J(X_{j_m(i,a)},A_{j_m(i,a)})\pi(a,\alpha)^2\right]\\
&=\frac{1}{N}\sum_{i=1}^{N}\hat{\sigma}^2_J(X_i,A_i)\pi(A_i,\alpha)^2+\frac{1}{N}\sum_{i=1}^{N}\frac{1}{M^2}\sum_{m=1}^{M}\sum_{a\neq A_i}\hat{\sigma}^2_J(X_{j_m(i,a)},A_{j_m(i,a)})\pi(a,\alpha)^2\\
&=\frac{1}{N}\sum_{i=1}^{N}\hat{\sigma}^2_J(X_i,A_i)\pi(A_i,\alpha)^2+\frac{1}{N}\frac{1}{M^2}K_M(i)\hat{\sigma}^2_J(X_i,A_i)\pi(A_i,\alpha)^2\\
&=\frac{1}{N}\sum_{i=1}^{N}\hat{\sigma}^2_J(X_i,A_i)\pi(A_i,\alpha)^2\Big(1+\frac{K_M(i)}{M^2}\Big)
\end{aligned}
$$

Combining the above we arrive at the estimator as in the main paper,

$$
\hat{V}^{\theta(\alpha|X)}_{M,J}=\frac{1}{N}\sum_{i=1}^{N}\left(\sum_{a\in\mathcal{A}(c)}\tilde{Y}_i(a)\pi(a,\alpha)-\hat{\theta}_M(\alpha)\right)^2+\frac{1}{N}\sum_{i=1}^{N}\Big(1+\frac{K_M(i)}{M^2}\Big)\pi(A_i,\alpha)^2\hat{\sigma}^2_J(X_i,A_i)
$$

APPENDIX C: TECHNICAL DETAILS FOR CHAPTER 4

C.1 Counterfactual probabilities

Some considerations for estimating the counterfactual probabilities $\omega(a, n, \alpha)$ are described below. All assumptions for identification discussed in the Section 4.4 of the main text are also made here; in particular that the ordering of individuals within clusters to be uninformative.

Let there be a random sample of $i = 1, \ldots, M$ clusters, and as in the main text denote by $O_i = \{N_i, L_i, A_i, Y_i\}$ the observed values of the random variables for cluster i. As described in Section 4.5 of the main text, $\widehat{\Pr}_\alpha(A_i = a|L_i, N_i)$ is calculated by substituting the estimates $(\hat{\gamma}_{0\alpha}, \hat{\beta}_1, \hat{\sigma})$ into the counterfactual cluster propensity score, $\Pr_\alpha(A_i = a|L_i, N_i)$. An estimator for the counterfactual probabilities is

$$\widetilde{\omega}(a, n, \alpha) = \left\{ \sum_{i=1}^{M} I(N_i = n) \right\}^{-1} \sum_{i=1}^{M} \widehat{\Pr}_\alpha(A_i = a|L_i, N_i),$$

which is not employed in the main text for the reasons described below.

Define $f(a) = \sum_{j=1}^{n} a_j$ to be the sum of the binary entries of $a \in \mathcal{A}(n)$. Letting $a, a' \in \mathcal{A}(n)$ be two vectors such that $f(a) = f(a')$, the assumed irrelevance of within-cluster ordering of individuals supposes that $\omega(a, n, \alpha) = \omega(a', n, \alpha)$. However, in any finite sample it is likely that $\widetilde{\omega}(a, n, \alpha) \neq \widetilde{\omega}(a', n, \alpha)$, which is an undesirable property of the above estimator. Thus, the estimator $\widetilde{\omega}(a, n, \alpha)$ is not pursued further here nor in the main text.

The method presented in Section 4.5 of the main text is discussed in further detail here. Under this assumption that the ordering of individuals within clusters to be uninformative, the counterfactual probabilities for clusters of size n and for a policy α

take on a maximum of $n+1$ unique values, rather than $2^n = |\mathcal{A}(n)|$. These counterfactual probabilities then arise from the strata of $\mathcal{A}(n,s) = \{a \in \mathcal{A}(n) | f(a) = s\}$ for $s = 0, 1, \ldots, n$, such that:

$$\omega(s, n, \alpha) = \sum_{a \in \mathcal{A}(n,s)} \omega(a, n, \alpha).$$

Thus for each $a \in \mathcal{A}(n)$ the counterfactual probabilities can be written as $\omega(a, n, \alpha) = \binom{n}{f(a)}^{-1} \omega(f(a), n, \alpha)$, and estimated by

$$\widehat{\omega}(a, n, \alpha) = \binom{n}{f(a)}^{-1} \widehat{\omega}(f(a), n, \alpha),$$

where $\widehat{\omega}(f(a), n, \alpha)$ is obtained from

$$\widehat{\omega}(f(a), n, \alpha) = \left\{ \sum_{i=1}^{M} I(N_i = n) \right\}^{-1} \sum_{a \in \mathcal{A}(n, f(a))} \sum_{i=1}^{M} \widehat{\Pr}_{\alpha}(A_i = a | L_i, N_i).$$

C.2 Simulating Data

For the simulation study in Section 4.7 of the main text, the true values of target estimands and nuisance causal parameters were determined empirically. The process is explained below. Recall the steps for generating a sample of data described in the main text: for each cluster i, step I was to generate the number N_i of individuals in the cluster, step II was to simulate covariates L_i, step III was to generate an observed treatment vector A_i, and step IV was to generate the observed outcome Y_i.

C.2.1 Determining the Counterfactual Model's Intercept

To determine $\gamma_{0\alpha}$ for $\alpha \in \{0.40, 0.50, 0.55, 0.75\}$, a grid of W-many potential values $\gamma_1^* < \gamma_2^* < \cdots < \gamma_w^* < \cdots < \gamma_W^*$ was proposed. For each $w = 1, \ldots, W$, the following

steps were carried out:

1. Steps I and II were repeated for $i = 1, \ldots, m_1 = 10^7$ clusters.

2. Treatment vectors were generated under policy α for m_1 clusters by replacing β_0 in step III with γ_w^*. That is, $A_{ij,w}$ for each individual j in cluster i was simulated from a Bernoulli distribution with probability $\mathcal{L}^{-1}(\gamma_w^* - 0.015L_{ij1} - 0.025L_{ij2} + b_i)$ where $b_i \sim N(0, 0.75)$.

3. The probability of obtaining treatment was assumed to equal the proportion of individuals in the dataset obtaining treatment, $p_w = \sum_{i=1}^{m_1} \sum_{j=1}^{N_i} I(A_{ij,w} = 1)/(\sum_{i=1}^{m_1} N_i)$.

For each α, $\gamma_{0\alpha}$ was determined to be the average of the γ_w^* that produced probabilities p_w closest to α, i.e., $\gamma_{0\alpha} = \text{mean}(\gamma_{w_l}^*, \gamma_{w_u}^*)$ where $w_l = \underset{\{w|p_w<\alpha\}}{\arg\max}(\alpha - p_w)$ and $w_u = \underset{\{w|p_w>\alpha\}}{\arg\max}(p_w - \alpha)$.

C.2.2 Determining Counterfactual Probabilities

For each α, $\omega(a, n, \alpha)$ was determined empirically from values of $\gamma_{0\alpha}$ determined as above. For each $n = 8, 22, 40$ and each α the following steps were carried out:

1. Step II was repeated for $i = 1, \ldots, m_2 = 10^8$ clusters of fixed size n.

2. For each cluster i a treatment vector was generated under each policy α by replacing β_0 in step III with the value $\gamma_{0\alpha}$ determined in Section C.2.1. That is, $A_{ij,\alpha}$ for each individual j in cluster i was simulated from a Bernoulli distribution with probability $\mathcal{L}^{-1}(\gamma_{0\alpha} - 0.015L_{ij1} - 0.025L_{ij2} + b_i)$ where $b_i \sim N(0, 0.75)$.

3. The counterfactual probabilities were defined as $\omega(a, n, \alpha) = \binom{n}{s}^{-1} \omega(s, n, \alpha)$ for each $s = 0, 1, \ldots, n$ where $\omega(s, n, \alpha) = m_2^{-1} \sum_{i=1}^{m_2} I\left(\sum_{j=1}^{n} A_{ij,\alpha} = s\right)$.

C.2.3 Simulating Potential Outcomes

For each $n = 8, 22, 40$ and each $s = 0, 1, \ldots, n$, let $a_{n,s}$ be the vector with s 1's followed by $(n - s)$ 0's. For each $n = 8, 22, 40$, the following steps were carried out:

1. Step II was repeated for $i = 1, \ldots, m_3 = 10^8$ clusters of fixed size n.

2. For each $s = 0, 1, \ldots, n$,

   (a) Individual potential outcomes $Y_{ij}(a_{n,s})$ were generated via the causal model analogous to the regression model specified in step IV for all individuals j in each cluster i. That is, $Y_{ij}(a_{n,s})$ was simulated from a Bernoulli distribution with mean $\Pr(Y_{ij}(a) = 1|L_{ij}) = \mathcal{L}^{-1}(0.1 - 0.05L_{ij1} + 0.5L_{ij2} - 0.5a_j + 0.2g(a_{-j}) - 0.25a_j g(a_{-j}))$, where $g(a_{-j}) = (N_i - 1)^{-1} \sum_{j' \neq j} a_{j'}$.

   (b) Then, $\overline{Y}_i(a_{n,s})$, $\overline{Y}_{0,i}(a_{n,s})$ and $\overline{Y}_{1,i}(a_{n,s})$ were computed for each cluster i according to their definitions presented in Section 4.3.1 of the main text.

   (c) Finally, define $\overline{\overline{Y}(a_{n,s})} = m_3^{-1} \sum_{i=1}^{m_3} \overline{Y}_i(a_{n,s})$ to be the average potential outcomes for all clusters when exposed to treatment $a_{n,s}$. For $t = 0, 1$ define $\overline{\overline{Y}_t(a_{n,s})} = m_3^{-1} \sum_{i=1}^{m_3} \overline{Y}_{t,i}(a_{n,s})$.

C.2.4 Determining Target Estimands

The values produced in Sections C.2.2 and C.2.3 were combined to determine the values of the target estimands. That is,

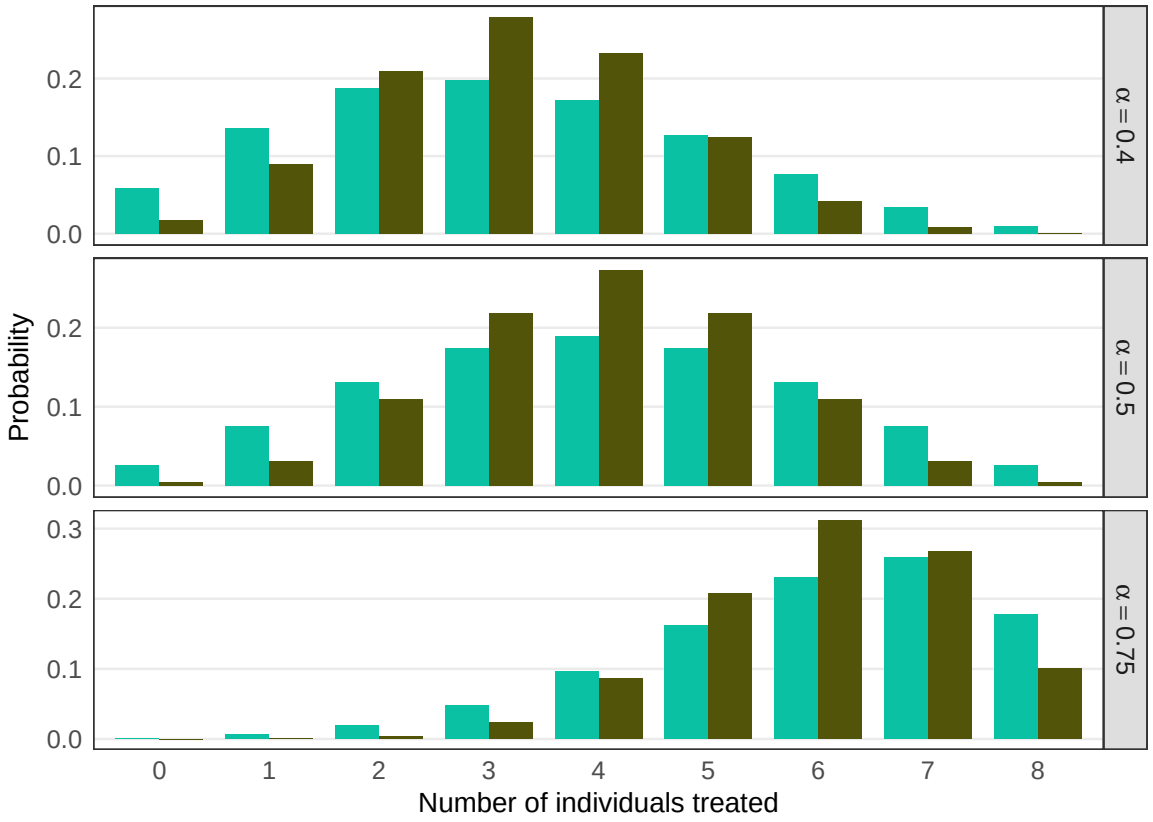
$$\mu(\alpha) = \sum_{n \in \{8,22,40\}} \left\{ \sum_{s=0}^{n} \left(\overline{\overline{Y}(a_{n,s})} \omega(s, n, \alpha) \right) \Pr(N_i = n) \right\},$$

and $\text{OE}(\alpha, \alpha') = \mu(\alpha) - \mu(\alpha')$. For $t = 0, 1$, $\text{SE}_t(\alpha, \alpha') = \mu_t(\alpha) - \mu_t(\alpha')$, where $\mu_t(\alpha) = \sum_{n \in \{8,22,40\}} \{ \sum_{s=0}^{n} (\overline{\overline{Y}_t(a_{n,s})} \omega(s, n, \alpha)) \Pr(N_i = n) \}$.

C.2.5 Empirical Comparison of Counterfactual Probabilities

Numerical differences in $\omega(a, n, \alpha)$ and $\omega_{\mathrm{B}}(a, n, \alpha)$ for the type B policies from Tchetgen Tchetgen and VanderWeele (2012) are dependent on the context and data generating process. Figure C.1 depicts the values of $\omega(s, n = 8, \alpha)$ determined in Section C.2.2 and the values of $\omega_{\mathrm{B}}(s, n = 8, \alpha) = \sum_{a \in \mathcal{A}(n,s)} \omega_{\mathrm{B}}(a, n = 8, \alpha)$. This figure illustrates the inequality $\omega(s, 8, \alpha) \neq \omega_{\mathrm{B}}(s, 8, \alpha)$ for all pairs of s and α for the data generating process described above.

Figure C.1: An empirical comparison of the counterfactual probabilities for the proposed estimands and the type B estimands for the data generating process in the simulation study described above and in the main text. The light green bars indicate $\omega(s, n, \alpha)$ and the dark brown bars indicate $\omega_{\mathrm{B}}(s, n, \alpha)$ for the type B policies from Tchetgen Tchetgen and VanderWeele (2012) for $s \in \{0, 1, \ldots, 8\}$, $n = 8$, and $\alpha \in \{0.4, 0.5, 0.75\}$.



The values of $\omega(s, n, \alpha)$ and $\omega_{\mathrm{B}}(s, n, \alpha)$ are particularly different when s is close to

0 or to n. For example, $\omega(0, 8, 0.40) = 0.059$ and $\omega_{\mathrm{B}}(0, 8, 0.40) = 0.017$, and so for the data generating process in this simulation study the proposed estimands confer $0.059/0.017 = 3.5$ times more weight to this category than the type B estimands.

REFERENCES

Abadie, A. and Imbens, G. W. (2006), "Large sample properties of matching estimators for average treatment effects," *Econometrica*, 74, 235–267.

— (2008), "On the failure of the bootstrap for matching estimators," *Econometrica*, 76, 1537–1557.

— (2016), "Matching on the estimated propensity score," *Econometrica*, 84, 781–807.

Ali, M., Emch, M. E., von Seidlein, L., Yunus, M., Sack, D. A., Rao, M., Holmgren, J., and Clemens, J. D. (2005), "Herd immunity conferred by killed oral cholera vaccines in Bangladesh: A reanalysis," *The Lancet*, 366, 44–49.

Ali, M., Emch, M. E., Yunus, M., and Clemens, J. D. (2009), "Modeling spatial heterogeneity of disease risk and evaluation of the impact of vaccination," *Vaccine*, 27, 3724–3729.

Ali, M., Kim, D. R., Kanungo, S., Sur, D., Manna, B., Digilio, L., Dutta, S., Marks, F., Bhattacharya, S. K., and Clemens, J. (2018), "Use of oral cholera vaccine as a vaccine probe to define the geographical dimensions of person-to-person transmission of cholera," *International Journal of Infectious Diseases*, 66, 90–95.

Aronow, P. M. (2012), "A general method for detecting interference between units in randomized experiments," *Sociological Methods & Research*, 41, 3–16.

Arpino, B., Benedictis, L. D., and Mattei, A. (2017), "Implementing propensity score matching with network data: the effect of the General Agreement on Tariffs and Trade on bilateral trade," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 66, 537–554.

Arpino, B. and Cannas, M. (2016), "Propensity score matching with clustered data. An application to the estimation of the impact of caesarean section on the Apgar score," *Statistics in Medicine*, 35, 2074–2091.

Arpino, B. and Mattei, A. (2016), "Assessing the causal effects of financial aids to firms in Tuscany allowing for interference," *The Annals of Applied Statistics*, 10, 1170–1194.

Arpino, B. and Mealli, F. (2011), "The specification of the propensity score in multilevel observational studies," *Computational Statistics & Data Analysis*, 55, 1770–1780.

Athey, S., Eckles, D., and Imbens, G. W. (2017), "Exact p-values for network interference," *Journal of the American Statistical Association*, 1–11.

Athey, S. and Imbens, G. W. (2016), "Recursive partitioning for heterogeneous causal effects," *Proceedings of the National Academy of Sciences*, 113, 7353–7360.

Austin, P. C. (2011), "An introduction to propensity score methods for reducing the effects of confounding in observational studies," *Multivariate behavioral research*, 46, 399–424.

Austin, P. C. and Stuart, E. A. (2015), "Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies," *Statistics in Medicine*, 34, 3661–3679.

Baird, S., Bohren, J. A., McIntosh, C., and Ozler, B. (2016), "Optimal Design of Experiments in the Presence of Interference," Working paper, URL `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2900967`.

Barkley, B. G. (2018), "clusteredinterference: Causal effects from observational studies with clustered interference," R package version 1.0.0. Available at `https://CRAN.R-project.org/package=clusteredinterference`.

Barkley, B. G., Hudgens, M. G., Clemens, J. D., Ali, M., and Emch, M. E. (2017), "Causal inference from observational studies with clustered interference," *arXiv preprint arXiv:1711.04834*.

Basse, G. and Feller, A. (2016), "Analyzing multilevel experiments in the presence of peer effects," *arXiv preprint arXiv:1608.06805*.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015), "Fitting Linear Mixed-Effects Models Using lme4," *Journal of Statistical Software*, 67, 1–48.

Bowers, J., Fredrickson, M. M., and Panagopoulos, C. (2013), "Reasoning about interference between units: A general framework," *Political Analysis*, 21, 97–124.

Brady, H. E. (2002), "Models of causal inference: Going beyond the Neyman-Rubin-Holland theory," in *Annual Meetings of the Political Methodology Group*.

Breslow, N. E. and Clayton, D. G. (1993), "Approximate inference in generalized linear mixed models," *Journal of the American Statistical Association*, 88, 9–25.

Brookhart, M. A., Schneeweiss, S., Rothman, K. J., Glynn, R. J., Avorn, J., and Stürmer, T. (2006), "Variable selection for propensity score models," *American Journal of Epidemiology*, 163, 1149–1156.

Brown Jr, B. W. (1980), "The crossover experiment for clinical trials," *Biometrics*, 69–79.

Busso, M., DiNardo, J., and McCrary, J. (2014), "New evidence on the finite sample properties of propensity score reweighting and matching estimators," *Review of Economics and Statistics*, 96, 885–897.

Cadarette, S. M., Gagne, J. J., Solomon, D. H., Katz, J. N., and Stürmer, T. (2010), "Confounder summary scores when comparing the effects of multiple drug exposures," *Pharmacoepidemiology and Drug Safety*, 19, 2–9.

Cannas, M., Arpino, B., and Billari, F. (2012), "Machine learning techniques for Propensity score matching with clustered data. A simulation study." in *46th Scientific Meeting of the Italian Statistical Society*.

Carnegie, N. B., Wang, R., and De Gruttola, V. (2016), "Estimation of the overall treatment effect in the presence of interference in cluster-randomized trials of infectious disease prevention," *Epidemiologic Methods*, 5, 57–68.

Cattaneo, M. D. (2010), "Efficient semiparametric estimation of multi-valued treatment effects under ignorability," *Journal of Econometrics*, 155, 138–154.

Centers for Disease Control and Prevention (CDC) (2012), "National, state, and local area vaccination coverage among children aged 19-35 months–United States, 2011," *MMWR Morb Mortal Wkly Rep*, 61, 689–696.

Chen, T., Zeng, D., and Wang, Y. (2015), "Multiple kernel learning with random effects for predicting longitudinal outcomes and data integration," *Biometrics*, 71, 918–928.

Clemens, J. D., Harris, J. R., Sack, D. A., Chakraborty, J., Ahmed, F., Stanton, B. F., Khan, M., Kay, B. A., Huda, N., Khan, M., et al. (1988), "Field trial of oral cholera vaccines in Bangladesh: results of one year of follow-up," *Journal of Infectious Diseases*, 158, 60–69.

Cochran, W. G. and Chambers, S. P. (1965), "The planning of observational studies of human populations," *Journal of the Royal Statistical Society. Series A (General)*, 128, 234–266.

Cole, S. R. and Frangakis, C. E. (2009), "The consistency statement in causal inference: a definition or an assumption?" *Epidemiology*, 20, 3–5.

Cole, S. R. and Hernán, M. A. (2004), "Adjusted survival curves with inverse probability weights," *Computer Methods and Programs in Biomedicine*, 75, 45–49.

Cox, D. R. (1958), *Planning of Experiments*, New York: John Wiley & Sons.

Dawid, A. P. (1979), "Conditional independence in statistical theory," *Journal of the Royal Statistical Society. Series B (Methodological)*, 1–31.

Dehejia, R. H. and Wahba, S. (1999), "Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs," *Journal of the American Statistical Association*, 94, 1053–1062.

— (2002), "Propensity score-matching methods for nonexperimental causal studies," *Review of Economics and Statistics*, 84, 151–161.

Drake, C. (1993), "Effects of misspecification of the propensity score on estimators of treatment effect," *Biometrics*, 1231–1236.

Eckles, D., Kizilcec, R. F., and Bakshy, E. (2016), "Estimating peer effects in networks with peer encouragement designs," *Proceedings of the National Academy of Sciences*, 113, 7316–7322.

Edwards, J. K., Cole, S. R., and Westreich, D. (2015), "All your data are always missing: incorporating bias due to measurement error into the potential outcomes framework," *International Journal of Epidemiology*.

Egger, P. H. and Von Ehrlich, M. (2013), "Generalized propensity scores for multiple continuous treatment variables," *Economics Letters*, 119, 32–34.

Feng, P., Zhou, X.-H., Zou, Q.-M., Fan, M.-Y., and Li, X.-S. (2012), "Generalized propensity score for estimating the average treatment effect of multiple treatments," *Statistics in Medicine*, 31, 681–697.

Fitzmaurice, G. M., Laird, N. M., and Ware, J. H. (2012), *Applied Longitudinal Analysis*, vol. 998, John Wiley & Sons.

Fokkema, M., Smits, N., Zeileis, A., Hothorn, T., and Kelderman, H. (2017), "Detecting treatment-subgroup interactions in clustered data with generalized linear mixed-effects model trees," *Behavior Research Methods*, 1–19.

Fong, C. and Imai, K. (2014), "Covariate balancing propensity score for general treatment regimes," *Princeton Manuscript*.

Forastiere, L., Airoldi, E. M., and Mealli, F. (2016a), "Identification and estimation of treatment and interference effects in observational studies on networks," *arXiv preprint arXiv:1609.06245*.

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

Foster, E. M. (2003), "Propensity score matching: an illustrative analysis of dose response," *Medical Care*, 41, 1183–1192.

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

Friedman, J. H. (1991), "Multivariate adaptive regression splines," *The Annals of Statistics*, 1–67.

Frölich, M. (2004a), "Finite-sample properties of propensity-score matching and weighting estimators," *Review of Economics and Statistics*, 86, 77–90.

— (2004b), "Programme evaluation with multiple treatments," *Journal of Economic Surveys*, 18, 181–224.

Gangl, M. (2010), "Causal inference in sociological research," *Annual Review of Sociology*, 36, 21–47.

Glymour, C. (1986), "Comment: Statistics and metaphysics," *Journal of the American Statistical Association*, 81, 964–966.

Graham, B. (2011), "Econometric methods for the analysis of assignment problems in the presence of complementarity and social spillovers," *Handbook of Social Economics*, 1, 965–1052.

Graham, D. J., McCoy, E. J., and Stephens, D. A. (2013), "Quantifying the effect of area deprivation on child pedestrian casualties by using longitudinal mixed models to adjust for confounding, interference and spatial dependence," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 176, 931–950.

Greenland, S., Pearl, J., and Robins, J. M. (1999a), "Causal diagrams for epidemiologic research," *Epidemiology*, 37–48.

Greenland, S. and Robins, J. M. (1986), "Identifiability, exchangeability, and epidemiological confounding," *International Journal of Epidemiology*, 15, 413–419.

Greenland, S., Robins, J. M., and Pearl, J. (1999b), "Confounding and collapsibility in causal inference," *Statistical Science*, 29–46.

Greenwood, M. and Yule, G. U. (1915), "The statistics of anti-typhoid and anti-cholera inoculations, and the interpretation of such statistics in general," *Proceedings of the Royal Society of Medicine*, 8, 113.

Grimes, D. A. and Schulz, K. F. (2002), "Bias and causal associations in observational research," *The Lancet*, 359, 248–252.

Grizzle, J. E. (1965), "The two-period change-over design and its use in clinical trials," *Biometrics*, 467–480.

Hajjem, A., Bellavance, F., and Larocque, D. (2011), "Mixed effects regression trees for clustered data," *Statistics & Probability Letters*, 81, 451–459.

— (2014), "Mixed-effects random forest for clustered data," *Journal of Statistical Computation and Simulation*, 84, 1313–1328.

Hajjem, A., Larocque, D., and Bellavance, F. (2017), "Generalized mixed effects regression trees," *Statistics & Probability Letters*, 126, 114–118.

Halloran, M. E., Haber, M., Longini, I. M., and Struchiner, C. J. (1991), "Direct and indirect effects in vaccine efficacy and effectiveness," *American Journal of Epidemiology*, 133, 323–331.

Halloran, M. E. and Struchiner, C. J. (1995), "Causal inference in infectious diseases," *Epidemiology*, 6, 142–151.

Hirano, K. and Imbens, G. W. (2004), "The propensity score with continuous treatments," *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives*, 226164, 73–84.

Hirano, K., Imbens, G. W., and Ridder, G. (2003), "Efficient estimation of average treatment effects using the estimated propensity score," *Econometrica*, 71, 1161–1189.

Ho, D. E., Imai, K., King, G., and Stuart, E. A. (2007), "Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference," *Political Analysis*, 15, 199–236.

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

Hong, G. and Raudenbush, S. W. (2006), "Evaluating kindergarten retention policy," *Journal of the American Statistical Association*, 101.

Horvitz, D. G. and Thompson, D. J. (1952), "A generalization of sampling without replacement from a finite universe," *Journal of the American Statistical Association*, 47, 663–685.

Huber, M., Lechner, M., and Wunsch, C. (2013), "The performance of estimators based on the propensity score," *Journal of Econometrics*, 175, 1–21.

Hudgens, M. G. and Halloran, M. E. (2008), "Toward causal inference with interference," *Journal of the American Statistical Association*, 103, 832–842.

Hulley, S., Grady, D., Bush, T., Furberg, C., Herrington, D., Riggs, B., Vittinghoff, E., Heart, Group, E. R. S. H. R., et al. (1998), "Randomized trial of estrogen plus progestin for secondary prevention of coronary heart disease in postmenopausal women," *JAMA*, 280, 605–613.

Iacus, S. M., King, G., Porro, G., and Katz, J. N. (2012), "Causal inference without balance checking: Coarsened exact matching," *Political Analysis*, 1–24.

Imai, K., King, G., and Stuart, E. A. (2008), "Misunderstandings between experimentalists and observationalists about causal inference," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 171, 481–502.

Imai, K. and Van Dyk, D. A. (2004), "Causal inference with general treatment regimes: Generalizing the propensity score," *Journal of the American Statistical Association*, 854–866.

Imbens, G. (2000), "The role of the propensity score in estimating dose-response functions," *Biometrika*, 87, 706.

Kang, H. and Imbens, G. (2016), "Peer encouragement designs in causal inference with partial interference and identification of local average network effects," *arXiv preprint arXiv:1609.04464*.

Kim, D. A., Hwong, A. R., Stafford, D., Hughes, D. A., O'Malley, A. J., Fowler, J. H., and Christakis, N. A. (2015), "Social network targeting to maximise population behaviour change: a cluster randomised controlled trial," *The Lancet*, 386, 145–153.

King, G. and Nielsen, R. (2016), "Why Propensity Scores Should Not Be Used for Matching," Working Paper, Institute for Quantitative Social Science, Harvard University, Boston.

Kluve, J., Schneider, H., Uhlendorff, A., and Zhao, Z. (2012), "Evaluating continuous training programmes by using the generalized propensity score," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 175, 587–617.

Koch, G. G., Amara, I. A., Stokes, M. E., and Gillings, D. B. (1980), "Some views on parametric and non-parametric analysis for repeated measurements and selected bibliography," *International Statistical Review/Revue Internationale de Statistique*, 249–265.

Kramer, A. D., Guillory, J. E., and Hancock, J. T. (2014), "Experimental evidence of massive-scale emotional contagion through social networks," *Proceedings of the National Academy of Sciences*, 111, 8788–8790.

Kreif, N., Grieve, R., Díaz, I., and Harrison, D. (2015), "Evaluation of the effect of a continuous treatment: A machine learning approach with an application to treatment for traumatic brain injury," *Health Economics*, 24, 1213–1228.

LaLonde, R. J. (1986), "Evaluating the econometric evaluations of training programs with experimental data," *The American Economic Review*, 76, 604–620.

Lechner, M. (2001), "Identification and estimation of causal effects of multiple treatments under the conditional independence assumption," in *Econometric Evaluation of Labour Market Policies*, Springer, pp. 43–58.

Lee, B. K., Lessler, J., and Stuart, E. A. (2010), "Improving propensity score weighting using machine learning," *Statistics in Medicine*, 29, 337–346.

Lewis, D. (1974), "Causation," *The Journal of Philosophy*, 70, 556–567.

— (2001), *Counterfactuals*, Blackwell.

Li, F., Zaslavsky, A. M., and Landrum, M. B. (2013), "Propensity score weighting with multilevel data," *Statistics in Medicine*, 32, 3373–3387.

Linden, A. and Yarnold, P. R. (2016), "Combining machine learning and propensity score weighting to estimate causal effects in multivalued treatments," *Journal of Evaluation in Clinical Practice*, 22, 871–881.

Little, R. J. and Yau, L. H. (1998), "Statistical techniques for analyzing data from prevention trials: Treatment of no-shows using Rubin's causal model." *Psychological Methods*, 3, 147.

Liu, L. and Hudgens, M. G. (2014), "Large sample randomization inference of causal effects in the presence of interference," *Journal of the American Statistical Association*, 109, 288–301.

Liu, L., Hudgens, M. G., and Becker-Dreps, S. (2016), "On inverse probability-weighted estimators in the presence of interference," *Biometrika*, 103, 829–842.

Luo, X., Small, D. S., Li, C.-S. R., and Rosenbaum, P. R. (2012), "Inference with interference between units in an fMRI experiment of motor inhibition," *Journal of the American Statistical Association*, 107, 530–541.

Manski, C. F. (2013), "Identification of treatment response with social interactions," *The Econometrics Journal*, 16, S1–S23.

— (2016), "Mandating vaccination with unknown indirect effects," *Journal of Public Economic Theory*.

McCaffrey, D. F., Griffin, B. A., Almirall, D., Slaughter, M. E., Ramchand, R., and Burgette, L. F. (2013), "A tutorial on propensity score estimation for multiple treatments using generalized boosted models," *Statistics in Medicine*, 32, 3388–3414.

Milborrow, S. (2014), "Notes on the earth package," URL `https://CRAN.R-project.org/package=earth`.

Morgan, S. L. and Winship, C. (2014), *Counterfactuals and Causal Inference*, Cambridge University Press.

Muñoz, I. D. and van der Laan, M. (2012), "Population intervention causal effects based on stochastic interventions," *Biometrics*, 68, 541–549.

Neyman, J. (1935), "Statistical problems in agricultural experimentation," *Supplement to the Journal of the Royal Statistical Society*, 2, 107–180.

Ogburn, E. L., VanderWeele, T. J., et al. (2014), "Causal diagrams for interference," *Statistical Science*, 29, 559–578.

Panozzo, C. A., Becker-Dreps, S., Pate, V., Funk, M. J., Stürmer, T., Weber, D. J., and Brookhart, M. A. (2013), "Patterns of rotavirus vaccine uptake and use in privately-insured US infants, 2006–2010," *PLoS One*, 8, e73825.

Panozzo, C. A., Becker-Dreps, S., Pate, V., Weber, D. J., Funk, M. J., Stürmer, T., and Brookhart, M. A. (2014), "Direct, indirect, total, and overall effectiveness of the rotavirus vaccines for the prevention of gastroenteritis hospitalizations in privately insured US children, 2007–2010," *American Journal of Epidemiology*, 179, 895–909.

Papadogeorgou, G., Mealli, F., and Zigler, C. (2017), "Causal inference for interfering units with cluster and population level treatment allocation programs," *arXiv preprint arXiv:1711.01280*.

Pearl, J. (1995), "Causal diagrams for empirical research," *Biometrika*, 82, 669–688.

— (1996), "Causation, action, and counterfactuals," in *Proceedings of the 6th conference on Theoretical aspects of rationality and knowledge*, Morgan Kaufmann Publishers Inc., pp. 51–73.

— (2010), "On the consistency rule in causal inference: Axiom, definition, assumption, or theorem?" *Epidemiology*, 21, 872–875.

Perez-Heydrich, C., Hudgens, M. G., Halloran, M. E., Clemens, J. D., Ali, M., and Emch, M. E. (2014), "Assessing effects of cholera vaccination in the presence of interference," *Biometrics*, 70, 731–741.

Pirracchio, R. and Carone, M. (2016), "The Balance Super Learner: A robust adaptation of the Super Learner to improve estimation of the average treatment effect in the treated based on propensity score matching," *Statistical Methods in Medical Research*.

Pirracchio, R., Petersen, M. L., and van der Laan, M. (2015), "Improving propensity score estimators' robustness to model misspecification using super learner," *American Journal of Epidemiology*, 181, 108–119.

Rassen, J. A., Shelat, A. A., Franklin, J. M., Glynn, R. J., Solomon, D. H., and Schneeweiss, S. (2013), "Matching by propensity score in cohort studies with three treatment groups," *Epidemiology*, 24, 401–409.

Richardson, T. S. and Robins, J. M. (2013), "Single world intervention graphs (SWIGs): A unification of the counterfactual and graphical approaches to causality," *Center for the Statistics and the Social Sciences, University of Washington Series. Working Paper*, 128, 2013.

Rigdon, J. (2015), "Package 'interferenceCI'," R package version 1.1. Available at `https://CRAN.R-project.org/package=interferenceCI`.

Rigdon, J. and Hudgens, M. G. (2015), "Exact confidence intervals in the presence of interference," *Statistics & Probability Letters*, 105, 130–135.

Robins, J. and Greenland, S. (1986), "The role of model selection in causal inference from nonexperimental data." *American Journal of Epidemiology*, 123, 392–402.

Robins, J. M. and Finkelstein, D. M. (2000), "Correcting for noncompliance and dependent censoring in an AIDS clinical trial with inverse probability of censoring weighted (IPCW) log-rank tests," *Biometrics*, 56, 779–788.

Robins, J. M., Hernan, M. A., and Brumback, B. (2000), "Marginal structural models and causal inference in epidemiology," *Epidemiology*, 11, 550–560.

Rosenbaum, P. R. (1987), "Model-based direct adjustment," *Journal of the American Statistical Association*, 82, 387–394.

— (2002), *Observational studies*, Springer.

Rosenbaum, P. R. and Rubin, D. B. (1983a), "Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome," *Journal of the Royal Statistical Society. Series B (Methodological)*, 45, 212–218.

— (1983b), "The central role of the propensity score in observational studies for causal effects," *Biometrika*, 70, 41–55.

— (1984), "Reducing bias in observational studies using subclassification on the propensity score," *Journal of the American Statistical Association*, 79, 516–524.

Ross, R. (1916), "An application of the theory of probabilities to the study of a priori pathometry. Part I," *Proceedings of the Royal Society of London. Series A, Containing papers of a mathematical and physical character*, 92, 204–230.

Rubin, D. B. (1974), "Estimating causal effects of treatments in randomized and nonrandomized studies." *Journal of Educational Psychology*, 66, 688–701.

— (1977), "Assignment to treatment group on the basis of a covariate," *Journal of Educational and Behavioral Statistics*, 2, 1–26.

— (1978), "Bayesian inference for causal effects: The role of randomization," *The Annals of Statistics*, 34–58.

— (1980), "Comment on "Randomization analysis of experimental data: The Fisher randomization test" by D. Basu," *Journal of the American Statistical Association*, 75, 591–593.

— (1990), "Comment: Neyman (1923) and causal inference in experiments and observational studies," *Statistical Science*, 5, 472–480.

— (1997), "Estimating causal effects from large data sets using propensity scores," *Annals of Internal Medicine*, 127, 757–763.

— (2005), "Causal inference using potential outcomes: Design, modeling, decisions," *Journal of the American Statistical Association*, 100, 322–331.

Rubin, D. B. and Thomas, N. (1996), "Matching using estimated propensity scores: relating theory to practice," *Biometrics*, 249–264.

Saul, B. C. (2017), "geex: An API for M-Estimation," R package version 1.0.3. Available at `https://CRAN.R-project.org/package=geex`.

Saul, B. C. and Hudgens, M. G. (2017), "A recipe for inferference: Start with causal inference. Add interference. Mix well with R." *Journal of Statistical Software*, 82, 1–21.

Saul, B. C. and Hudgens, M. G. (2017), "The calculus of M-estimation in R with geex," *ArXiv e-prints arXiv:1709.01413*.

Schisterman, E. F., Cole, S. R., and Platt, R. W. (2009), "Overadjustment bias and unnecessary adjustment in epidemiologic studies," *Epidemiology (Cambridge, Mass.)*, 20, 488.

Schneeweiss, S., Rassen, J. A., Glynn, R. J., Avorn, J., Mogun, H., and Brookhart, M. A. (2009), "High-dimensional propensity score adjustment in studies of treatment effects using health care claims data," *Epidemiology (Cambridge, Mass.)*, 20, 512.

Schuler, M. S., Chu, W., and Coffman, D. (2016), "Propensity score weighting for a continuous exposure with multilevel data," *Health Services and Outcomes Research Methodology*, 16, 271–292.

Schulz, K. F. and Grimes, D. A. (2002), "Generation of allocation sequences in randomised trials: Chance, not choice," *The Lancet*, 359, 515–519.

Sekhon, J. S. (2008), "The Neyman-Rubin model of causal inference and estimation via matching methods," *The Oxford Handbook of Molitical Methodology*, 2.

— (2009), "Opiates for the matches: Matching methods for causal inference," *Annual Review of Political Science*, 12, 487–508.

Setoguchi, S., Schneeweiss, S., Brookhart, M. A., Glynn, R. J., and Cook, E. F. (2008), "Evaluating uses of data mining techniques in propensity score estimation: A simulation study," *Pharmacoepidemiology and Drug Safety*, 17, 546–555.

Smith, J. A. and Todd, P. E. (2005), "Does matching overcome LaLonde's critique of nonexperimental estimators?" *Journal of Econometrics*, 125, 305–353.

Sobel, M. E. (2006), "What do randomized studies of housing mobility demonstrate? Causal inference in the face of interference," *Journal of the American Statistical Association*, 101, 1398–1407.

Sofrygin, O. and van der Laan, M. J. (2017), "Semi-parametric estimation and inference for the mean outcome of the single time-point intervention in a causally connected population," *Journal of Causal Inference*, 5.

Spreeuwenberg, M. D., Bartak, A., Croon, M. A., Hagenaars, J. A., Busschbach, J. J., Andrea, H., Twisk, J., and Stijnen, T. (2010), "The multiple propensity score as control for bias in the comparison of more than two treatment arms: An introduction from a case study in mental health," *Medical Care*, 166–174.

Stefanski, L. A. and Boos, D. D. (2002), "The calculus of M-estimation," *The American Statistician*, 56, 29–38.

Stuart, E. A. (2010), "Matching methods for causal inference: A review and a look forward," *Statistical Science*, 25, 1.

Tate, J. E., Panozzo, C. A., Payne, D. C., Patel, M. M., Cortese, M. M., Fowlkes, A. L., and Parashar, U. D. (2009), "Decline and change in seasonality of US rotavirus activity after the introduction of rotavirus vaccine," *Pediatrics*, 124, 465–471.

Tchetgen Tchetgen, E. J. and VanderWeele, T. J. (2012), "On causal inference in the presence of interference," *Statistical Methods in Medical Research*, 21, 55–75.

Therneau, T. M., Atkinson, E. J., et al. (1997), "An introduction to recursive partitioning using the RPART routines," Technical Report 61. URL `http://www.mayo.edu/hsr/techrpt/61.pdf`.

Toulis, P. and Kao, E. K. (2013), "Estimation of causal peer influence effects," in *International Conference on Machine Learning*, pp. 1489–1497.

Ugander, J., Karrer, B., Backstrom, L., and Kleinberg, J. (2013), "Graph cluster randomization: Network exposure to multiple universes," in *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, pp. 329–337.

van der Laan, M. J. (2014), "Causal inference for a population of causally connected units," *Journal of Causal Inference*, 2, 13–74.

VanderWeele, T. J. (2009), "Concerning the consistency assumption in causal inference," *Epidemiology*, 20, 880–883.

VanderWeele, T. J. and An, W. (2013), "Social networks and causal inference," in *Handbook of Causal Analysis for Social Research*, Springer, pp. 353–374.

VanderWeele, T. J. and Shpitser, I. (2011), "A new criterion for confounder selection," *Biometrics*, 67, 1406–1413.

VanderWeele, T. J., Tchetgen, E. J. T., and Halloran, M. E. (2014), "Interference and sensitivity analysis," *Statistical Science*, 29, 687–706.

VanderWeele, T. J. and Tchetgen Tchetgen, E. J. (2011a), "Bounding the infectiousness effect in vaccine trials," *Epidemiology (Cambridge, Mass.)*, 22, 686–693.

— (2011b), "Effect partitioning under interference in two-stage randomized vaccine trials," *Statistics & Probability Letters*, 81, 861–869.

Vansteelandt, S., Bekaert, M., and Claeskens, G. (2012), "On model selection and model misspecification in causal inference," *Statistical Methods in Medical Research*, 21, 7–30.

Verbitsky-Savitz, N. and Raudenbush, S. W. (2012), "Causal inference under interference in spatial settings: A case study evaluating community policing program in Chicago," *Epidemiologic Methods*, 1.

Verhave, T., Owen, J. E., and Robbins, E. B. (1959), "The effect of morphine sulfate on avoidance and escape behavior," *Journal of Pharmacology and Experimental Therapeutics*, 125, 248–251.

Watkins, S., Jonsson-Funk, M., Brookhart, M. A., Rosenberg, S. A., O'Shea, T. M., and Daniels, J. (2013), "An Empirical Comparison of Tree-Based Methods for Propensity Score Estimation," *Health Services Research*, 48, 1798–1817.

Westreich, D. (2017), "From Patients to Policy: Population Intervention Effects in Epidemiology," *Epidemiology*, 28, 525–528.

Westreich, D., Cole, S. R., Funk, M. J., Brookhart, M. A., and Stürmer, T. (2011), "The role of the c-statistic in variable selection for propensity score models," *Pharmacoepidemiology and Drug Safety*, 20, 317–320.

Westreich, D., Lessler, J., and Funk, M. J. (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, 826–833.

Woo, M.-J., Reiter, J. P., and Karr, A. F. (2008), "Estimation of propensity scores using generalized additive models," *Statistics in Medicine*, 27, 3805–3816.

Wyss, R., Ellis, A. R., Brookhart, M. A., Girman, C. J., Funk, M. J., LoCasale, R., and Stürmer, T. (2014), "The role of prediction modeling in propensity score estimation: an evaluation of logistic regression, bCART, and the covariate-balancing propensity score," *American Journal of Epidemiology*, 180, 645–655.

Yang, S. (2017), "Propensity score weighting for causal inference with clustered data," *arXiv preprint arXiv:1703.06086*.

Yang, S., Imbens, G. W., Cui, Z., Faries, D. E., and Kadziola, Z. (2016), "Propensity score matching and subclassification in observational studies with multi-level treatments," *Biometrics*, 72, 1055–1065.

Yang, W., Joffe, M. M., Hennessy, S., and Feldman, H. I. (2014), "Covariance adjustment on propensity parameters for continuous treatment in linear models," *Statistics in Medicine*, 33, 4577–4589.

Zanutto, E., Lu, B., and Hornik, R. (2005), "Using propensity score subclassification for multiple treatment doses to evaluate a national antidrug media campaign," *Journal of Educational and Behavioral Statistics*, 30, 59–73.

Zhang, B., Tsiatis, A. A., Laber, E. B., and Davidian, M. (2012), "A robust method for estimating optimal treatment regimes," *Biometrics*, 68, 1010–1018.

Zhao, Y., Zeng, D., Rush, A. J., and Kosorok, M. R. (2012), "Estimating individualized treatment rules using outcome weighted learning," *Journal of the American Statistical Association*, 107, 1106–1118.

Zhu, Y., Coffman, D. L., and Ghosh, D. (2015), "A boosting algorithm for estimating generalized propensity scores with continuous treatments," *Journal of Causal Inference*, 3, 25–40.

Zigler, C. M., Dominici, F., and Wang, Y. (2012), "Estimating causal effects of air quality regulations using principal stratification for spatially correlated multivariate intermediate outcomes," *Biostatistics*, 13, 289–302.