# CAUSAL INFERENCE FOR BINARY DATA WITH INTERFERENCE

Joseph Rigdon

A dissertation submitted to the faculty at 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.

Chapel Hill
2015

Approved by:

Michael Hudgens

Michael Emch

Amy Herring

Gary Koch

Mark Weaver

## ABSTRACT

Joseph Rigdon: Causal Inference for Binary Data with Interference
(Under the direction of Michael Hudgens)

Developing methods to quantify the effects of interventions to prevent infectious diseases in the presence of interference is the overall objective of this research. Interference is present when an individual's outcome is affected by the treatment of any other individuals under study.

First, two methods are developed for constructing randomization based confidence intervals for the average effect of a treatment on a binary outcome without interference. The methods are nonparametric and require no assumptions about random sampling from a larger population. Both of the resulting $1-\alpha$ confidence intervals are exact and guaranteed to have width no greater than one. In contrast, previously proposed asymptotic confidence intervals are not exact and may have width greater than one. The first approach combines Bonferroni adjusted prediction intervals for the attributable effects in the treated and untreated. The second method entails inverting a permutation test. While simulations show that the permutation based confidence intervals have smaller average width, the attributable effects based confidence intervals are more computationally feasible as sample size increases. Extensions that allow for stratifying on categorical baseline covariates are also discussed.

Secondly, for a two-stage randomized experiment assuming stratified interference, methods are developed for constructing exact confidence intervals for the direct, indirect, total, and overall effect of a treatment on a binary outcome. The methods are nonparametric and require no assumptions about random sampling from a larger population. The new exact confidence intervals are compared via simulation with previously proposed exact and asymptotic confidence intervals. While the asymptotic intervals do not maintain nominal coverage for

certain simulation setups, the new exact confidence intervals maintain nominal coverage for all setups and have narrower width than the previously proposed exact confidence interval.

Thirdly, we consider a Bayesian approach to causal inference with interference in an observational study under the assumption that the treatment assignment mechanism is ignorable. We compare the methods via a simulation study to previously proposed IPW estimators. The methods are applied to data from the 2007 Demographic and Health Survey in the Democratic Republic of the Congo, examining the impact of individual and community bed net use on malaria.

## ACKNOWLEDGMENTS

I want to thank many people who took this journey with me. First, I'd like to thank Michael Hudgens. Michael took a chance on me as one of his first PhD students when few others would. He provided the constructive criticism I needed to grow as a statistician and researcher. His expectations pushed me to a level I could not have achieved on my own. Michael is a brilliant scholar who is quietly leaving a legacy of excellence one paper at a time. Second, I'd like to thank the members of my committee. Amy Herring's expertise on Bayesian modeling was invaluable. Gary Koch's financial support for JSM in 2014 allowed me to find a job! Mark Weaver showed me that a collaborative career in a medical school was something I will seriously consider. Mike Emch provided us with a very interesting data set and expanded my knowledge of malaria and geographical health patterns in Africa.

Finally, I'd like to thank my family and friends – without your love and support, who knows where I would be.

## TABLE OF CONTENTS

## LIST OF TABLES

## LIST OF FIGURES

# CHAPTER 1: INTRODUCTION AND LITERATURE REVIEW

## 1.1 Introduction

Developing methods to quantify the effects of interventions to prevent infectious diseases in the presence of interference is the overall objective of this research. In a randomized experiment or observational study, interference is present when an individual's outcome is affected by the exposure(s) of any other individual(s) under study. Interference is a phenomenon that appears in many fields of study, e.g., infectious diseases, education, sociology, and various other settings in which individuals communicate or compete. A canonical example of interference occurs when vaccinated individuals can protect unvaccinated individuals from disease when all individuals are in close contact with one another. Standard statistical methods may indicate that the vaccinated individuals in such a situation are minimally less likely to develop disease than the unvaccinated, thus leading an investigator to believe that vaccine is ineffective. This is a possibly erroneous conclusion, as it is plausible that the vaccinated individuals are protecting the group at large from disease, i.e., herd immunity. If interference is not accounted for in such a situation, a potentially useful vaccine could be deemed ineffective. Therefore, it is of great public health importance to properly evaluate vaccines in the presence of interference.

Building upon ideas in Halloran et al. (1991), Hudgens and Halloran (2008) defined four treatment effects of interest in the presence of interference for a finite population using the potential outcomes framework. The direct effect of treatment is the difference in outcomes when vaccinated versus not at a fixed level of group treatment coverage. The indirect effect of treatment is the difference in outcomes when unvaccinated in a low treatment coverage group versus unvaccinated in a high treatment coverage group. The indirect effect measures the "spillover" effects of vaccine. The total causal effect is the difference in outcomes when

unvaccinated in a low treatment coverage group versus vaccinated in a high treatment coverage group. The overall effect of treatment is the difference in outcomes in a low treatment coverage group versus a high treatment coverage group. The focus of this research is to develop novel statistical methods to draw inference about the direct, indirect, total and overall effect of a treatment in the presence of interference.

To accomplish these goals, we 1) develop exact confidence intervals for the direct effect of treatment on a binary outcome when interference is not present in a randomized trial with one or more groups, then 2) develop exact confidence intervals for the four treatment effects on a binary outcome in a randomized trial when interference is present, and conclude with 3) building methods for inference on the four effects in a non-randomized real setting in which interference may be present.

## 1.2 Motivating Examples

Oral cholera vaccines were administered to residential areas, called baris, in Bangladesh in an individually randomized placebo-controlled trial (Ali et al., 2005). The baris had varying levels of proportion of residents who received vaccine, or vaccine coverage, due to differing levels of participation in the trial. Summary measures of the incidence of cholera during the first year of follow-up of the trial are displayed in Table 1.1.

Table 1.1: Risk of cholera in recipients of killed oral cholera vaccines of placebo, by level of coverage of the bari during one year of follow-up

| | | Vaccine recipients | | | Placebo recipients | | |
|---|---|---|---|---|---|---|---|
| Coverage | Population | Total | Cases | Risk per 1,000 | Total | Cases | Risk per 1,000 |
| >50% | 22,394 | 12,541 | 16 | 1.27 | 6,082 | 9 | 1.47 |
| 41-50% | 24,159 | 11,513 | 26 | 2.26 | 5,801 | 27 | 4.65 |
| 36-40% | 24,583 | 10,772 | 17 | 1.58 | 5,503 | 26 | 4.72 |
| 28-35% | 25,059 | 8,883 | 22 | 2.48 | 4,429 | 26 | 5.87 |
| <28% | 24,954 | 5,627 | 15 | 2.66 | 2,852 | 20 | 7.01 |

In Table 1.1, placebo and vaccine recipients in the >50% coverage group had nearly the same risk of cholera, yet placebo recipients in the <28% coverage group werre nearly three

times as likely to experience cholera as vaccine recipients. When more individuals in a group were vaccinated, the outcome of a single individual, regardless of vaccine status, is more likely to be no cholera. Thus, interference may be present, as an individual's cholera outcome is affected by the vaccine status of other members of the group.

In addition to cholera, malaria is another infectious disease setting where interference is found. Treated and untreated bed nets are a frequently used intervention to protect individuals in at-risk areas from mosquitos infected with malaria. Empirical evidence indicates that as the bed net coverage level goes up in a group of individuals, malaria incidence tends to go down (Hii et al., 2001; Hawley et al., 2003; Messina et al., 2011). In the Wosera area of Papua New Guinea, Hii et al. (2001) found that untreated bed nets have a substantial impact on malaria prevalence in high coverage areas that is greater than can be accounted for by personal protection. In the Nyanza Province in western Kenya, Hawley et al. (2003) found a similar strong community effect of treated bed nets, hypothesizing that these effects were due to reduced longevity of mosquito populations forced to expend extra energy in search of human blood. Using data collected in the 2007 Demographic and Health Survey (DHS) in the Democratic Republic of the Congo (DRC), Messina et al. (2011) modeled malaria status as a function of group and individual covariates in a multilevel logistic regression. Individual bed net usage was significant when entered alone in the model, but was no longer significant when group level bed net coverage was included in the model, suggesting that herd immunity may be occurring within communities. In the final section of this dissertation, we develop methods for inference on the four effects of interest in a re-analysis of the 2007 DHS data.

## 1.3 Causal Inference

The fundamental goal of many public health studies is draw inferences on the cause of a health-related outcome. As Gelman (2011) states, there are two broad classes of inferential questions to consider in a causal analysis. First, there are forward inferential questions that follow individuals with a treatment or exposure and monitor outcomes. Does a novel treatment lead to longer survival time? If I study more, will I score higher on the test? Second,

there are reverse inferential questions that start with an outcome in mind and try to identify its cause. Why did a certain group of students perform best on the SAT? Why did the basketball team win the title? This dissertation will be concerned with questions belonging to the first broad category.

Randomized experiments are the ideal means through which investigators can answer forward inferential questions about causality. The act of randomizing treatments leads to unbiased estimators of treatment effects and facilitates the use of simple and conventional statistical methods. Going beyond conventional methods, the paradigm of potential outcomes offers intuitive and transparent statements about causality in a randomized experiment (Little and Rubin, 2000). Extending the ideas of Neyman (1923), Rubin (1974) developed a formal framework defining causal effects as comparisons of potential outcomes that would have been observed under different exposures to treatments. In this view, each individual's potential outcome on control serves as his or her own control. Conceptual clarity is provided as the effects of causes, which can be precisely defined using potential outcomes, rather than the causes of effects are examined (Holland, 1986). As only one outcome is observed per unit, statistical methods are necessary to make statements about the unobserved potential outcomes for each unit and thus the causal effects. Greenland (2000) outlined the benefits of such an approach, also called the "counterfactual" approach. In addition to conceptual clarity, this approach allows for extensions to more complex settings than randomized experiments such as observational studies or settings with interference. We will make use of this paradigm to accomplish all three objectives stated in the introduction.

Consider a study with $i = 1, ..., k$ groups of size $n_i$ such that there are $n = \sum_{i=1}^{k} n_i$ total individuals in the study. For individual $j$ in group $i$, let $Y_{ij}$ represent the observed outcome and $Z_{ij}$ represent the observed treatment or exposure. Depending on the context of the investigation, $Y_{ij}$ may be binary, ordinal, or continuous. Let $Z_{ij}$ be dichotomous where $Z_{ij} = 0$ indicates no treatment and $Z_{ij} = 1$ indicates treatment.

In the potential outcomes paradigm, $Y_{ij}$ is one of many potential outcomes in the set $\{y_{ij}(z) : z \in \mathcal{Z}\}$ that could have been revealed by the random variable $Z_{ij}$. Thus, $Y_{ij}$ is

also a random variable. The set $\{y_{ij}(z) : z \in \mathcal{Z}\}$ could contain a unique potential outcome for each of the $2^n$ assignments of treatment or no treatment for the $n$ individuals in the study, but typically the number of potential outcomes per individual are limited by assumptions about interference and study design. Additionally, the potential outcome $y_{ij}(z)$ could be viewed as fixed or random depending on the type of inference being used. From the finite population perspective, $y_{ij}(z)$ is fixed and inferential targets are functions of the fixed potential outcome structure, e.g., the sample average causal effect in group $i$ without interference $n^{-1}\sum_{j=1}^{n_i}\{y_{ij}(0) - y_{ij}(1)\}$. From the superpopulation perspective, $y_{ij}(z)$ is considered a random variable and inferential targets are functions of its distribution, e.g., the population average causal effect in group $i$ without interference $E[y_i(0) - y_i(1)]$. In this section, the finite population perspective is adopted, but in later sections, the superpopulation perspective will be employed.

In the simplest case with no interference, $\{y_{ij}(z) : z \in \mathcal{Z}\}$ contains $y_{ij}(0)$, the potential outcome on control, and $y_{ij}(1)$, the potential outcome on treatment. In this case, a natural causal estimand is the direct effect of vaccine, $y_{ij}(0) - y_{ij}(1)$. Supposing the potential outcomes were binary, if the direct effect in individual $j$ were 1, it could be concluded that vaccine prevents disease; however, both potential outcomes are never observed together and therein lies the necessity for statistical methods.

As Rubin (2005) notes, causal inference is impossible without assumptions and it is the scientific quality of those assumptions, not their existence, that is critical. A standard assumption in the potential outcomes framework is the stable unit treatment value assumption (SUTVA) (Rubin, 1980) that assumes 1) there are not multiple forms of treatment and 2) there is no interference between units (Cox, 1958). In this dissertation, we consider problems where assumption 2) is violated, i.e., the set $\{y_{ij}(z) : z \in \mathcal{Z}\}$ contains elements beyond $y_{ij}(0)$ and $y_{ij}(1)$.

## 1.4 Randomization Based Inference

Any randomized experiment will only reveal one potential outcome for each individual, necessitating statistical methods to draw inferences about the unobserved potential outcomes. Randomization based inference can be employed to make inferences about unobserved quantities. Dating back to Fisher (1935), this mode of inference uses the randomization process in the experimental design to induce the distribution of a test statistic under a null hypothesis. As no distributional assumptions are needed for the data generating process, no likelihood function is formed. No assumptions are needed about sampling from a superpopulation as the target of inference is the study population itself. A classic example of randomization based inference is Fisher's Exact Test for binary exposure and outcome data. The distribution of any cell in the $2 \times 2$ table formed by the observed data is known be hypergeometric under the sharp null hypothesis that $y_j(0) = y_j(1)$ for individuals $j = 1, ..., n$.

Rubin (1974) delineated two formal benefits of randomization in an experiment in which $n$ of $2n$ subjects are randomized to treatment with no interference. First, the observed difference in the mean of the outcomes of the $n$ subjects who receive control minus the mean of the outcomes of the $n$ subjects who receive treatment is an unbiased estimator of the average causal effect $\sum_{j=1}^{2n}\{y_j(0) - y_j(1)\}$ (Neyman, 1923). Second, precise probabilistic statements can be made indicating how likely the observed difference in treatment group means is under a "sharp" null hypothesis. Hypotheses tested must be "sharp" in the sense that they specify all possible potential outcomes. The validity of a hypothesis is tested by re-running the experiment, i.e., by re-randomizing individuals to treatment or control. A test statistic is computed under each re-randomization including the observed randomization. If the observed data test statistic is extreme relative to the test statistics generated by the re-randomizations under the null hypothesis, then there is evidence that the hypothesis is false.

Stewart (2002) likens randomization based inference to the movie Groundhog Day, in which Bill Murray's character re-lives the same day over and over again. The act of re-living the same day can be thought of as a re-randomization of treatments. Eventually, after enough iterations of this process, he is able to infer the true structure of his environment and live the

perfect day. In the vaccine trial setting, this would be analogous to revealing all potential outcomes and assigning the most beneficial treatment to each individual. The analogy can be taken a step further in that any inferences that Bill Murray's character makes in Groundhog Day only apply to his environment, and that any inferences made in the vaccine trial only apply to the individuals in the study.

To illustrate these points, consider the following example. Two of four participants are to be randomized to vaccine. The outcome of interest is the number of visits to the doctor in the year after vaccine is administered. Assuming no interference is present, each of the $j = 1, ..., 4$ individuals has two potential outcomes: $y_j(0)$, the number of visits on control, and $y_j(1)$, the number of visits on vaccine. Let $\tau_j = y_j(0) - y_j(1)$ be the causal effect of vaccine for individual $j$. Table 1.2 displays all potential outcomes, half of which can be revealed by investigation, such that the true average causal effect is $\tau = (1/4)\sum_{j=1}^{4} \tau_j = 4$.

Table 1.2: Vaccine trial example with number of visits to doctor in the following year as outcome

| Individual | $y(0)$ | $y(1)$ | $\tau_j$ |
|---|---|---|---|
| 1 | 7 | 4 | 3 |
| 2 | 10 | 5 | 5 |
| 3 | 5 | 1 | 4 |
| 4 | 4 | 0 | 4 |

Let $Z_j = 1$ if randomized to vaccine and 0 if placebo. Let the estimator be $\hat{\tau} = (1/2)\sum_{j=1}^{4}(1 - Z_j)y_j(0) - (1/2)\sum_{j=1}^{4} Z_j y_j(1)$. There are $\binom{4}{2}$ equals 6 ways of randomizing two of four individuals to vaccine, each occurring with equal probability 1/6. The 6 possibilities of the vector of treatments $Z$ with corresponding estimated average causal effect $\hat{\tau}$ are 1100 and 0, 1010 and 4.5, 1001 and 5.5, 0110 and 2.5, 0101 and 3.5, and 0011 and 8. The estimator $\hat{\tau}$ is unbiased as $E[\hat{\tau}] = (1/6)(0 + 4.5 + 5.5 + 2.5 + 3.5 + 8) = 4$.

To illustrate the second point made by Rubin (1974) about the randomization-based p-value, let the observed data be $Z = 0101$ and $\hat{\tau} = 3.5$. As the randomization process only reveals one potential outcome for each individual, $y_1(1)$, $y_2(0)$, $y_3(1)$, and $y_4(0)$ remain unknown. Under the sharp null hypothesis $H^0 : \tau_j = \tau_j^0$ for each $j$, the blanks can be filled

in. If the missing outcome for individual $j$ is $y_j(0)$, it is known under the null as $y_j(1) + \tau_j^0$, and if the missing outcome is $y_j(1)$, it is known under the null as $y_j(0) - \tau_j^0$. For illustration, consider Fisher's sharp null that $\tau_j = 0$ for all $j$. Under this null, the 6 possibilities of the vector $Z$ with corresponding estimated average causal effect $\hat{\tau}$ are 1100 and -3.5, 1010 and -3.5, 1001 and 1.5, 0110 and 1.5, 0101 and 3.5 (the observed data), and 0011 and 3.5. A measure of extremeness of the observed $\hat{\tau}$ is the proportion of re-randomized $\hat{\tau}$ values are farther in absolute distance from the hypothesized center of the distribution, or the permutation test p-value $p = (1/6)\sum_{c=1}^{6} 1\{|\hat{\tau}^c - 0| \geq |\hat{\tau} - 0|\}$, where $\hat{\tau}^c$ denotes the value of $\hat{\tau}$ under re-randomization $c = 1, ..., 6$. If $p$ is large, then the observed $\hat{\tau}$ is close to the center of the distribution of $\hat{\tau}$ under the null, lending support to the null being true. If $p$ is small, then the observed $\hat{\tau}$ is far from the center of the distribution of $\hat{\tau}$ under the null, lending support to the null being false. In this example, $p = 4/6$ lends support to the null being true.

## 1.5 Incorporating Covariate Information

Although the estimator $\hat{\tau}$ is unbiased, it can be highly variable. For the data in Table 1.2, the estimator $\hat{\tau}$ can take on a value as large as 8 even though the true $\tau$ is equal to 4 in. Matching units that would respond similarly to vaccine can lower the variance of $\tau$ (Rubin, 1974). To illustrate this point, again consider the data in Table 1.2. Individuals 1 and 2 have similar outcomes on vaccine and placebo and are "matched" in some sense, as are individuals 3 and 4. If randomization is limited so that one individual gets vaccine among 1 and 2 and only one individual gets vaccine among 3 and 4, the 4 possible values of $Z$ and $\hat{\tau}$ are 1010 and 4.5, 1001 and 5.5, 0110 and 2.5, and 0101 and 3.5. The variance of $\hat{\tau}$ has been lowered from 7.4 to 1.7, while $\hat{\tau}$ remains unbiased.

This process of matching is straightforward in the physical sciences, where for example it is easy to apply treatment 1 and treatment 2 to two blocks of wood cut from the same tree. Matching can become tricky when dealing with humans and social science phenomena. When many categorical and continuous covariates are collected as in most public health studies, alternatives to matching include covariance adjustment, propensity score matching,

and inverse probability weighting.

### 1.5.1 Covariance Adjustment

Rosenbaum (2002a) maps out a strategy for covariance adjustment based on important covariates for a continuous outcome. Let $y(0)$ be the vector of potential outcomes on control, $y(1)$ on treatment, $Z$ the vector of treatment assignments such that $m$ of $n$ receive vaccine, $\delta$ the vector of causal effects such that $\tau = 1'\delta/n$, and $Y = Z'y(1) + (1-Z)'y(0)$, the vector of observed outcomes. Rosenbaum (2002a) assumes additivity in the sense that only hypotheses of the form $H^0 : \delta = \tau^0$ are considered, where $\tau^0$ is a scalar constant. Let $q$ be the vector of ranks of $y(0) = Y + Z\tau^0$ under the null. Then, the plausibility of $H^0$ can be tested by comparing the Wilcoxon rank-sum statistic, $q'Z$, to its distribution under the null, a sum of $m$ numbers randomly selected from $\{1, ..., n\}$, a unimodal distribution with mean $m(n+1)/2$. The p-value is the proportion of these sums that are more extreme than the observed $q'Z$, i.e., that they are farther in way from $m(n+1)/2$ in absolute distance. A $1-\alpha$ confidence interval for $\tau$ assuming additivity includes all $\tau^0$ such that this p-value is greater than or equal to $\alpha$.

Assuming additivity, Rosenbaum (2002a) refines this process to include covariates. Let $X$ be a matrix of measured covariates known to be associated with $Y$, but not including $Z$. Under $H^0 : \delta = \tau^0$, let $\tilde{\epsilon}(Y + Z\tau^0) = \tilde{\epsilon}(y(0)) = e$ be the residuals from some model of $y(0)$ as a function of $X$. Let $q_e$ be the ranks of $e$. Under the null $H^0 : \delta = \tau^0$, the quantities $y(0)$ and $X$ are fixed as they do not depend on $Z$. Thus, under the null, the Wilcoxon rank-sum statistic, $q_e'Z$, follows the same distribution described in the previous paragraph. A p-value can be found to test the plausibility of this hypothesis in the same manner as before, and a $1-\alpha$ confidence interval for $\tau$ assuming additivity follows analogously. This procedure eliminates some of the variation in responses due to the covariates as $e$ is expected to be less dispersed than $y(0)$.

### 1.5.2 Propensity Score Matching

In a non-randomized study, e.g. an observational study, important covariates may not be balanced across the vaccine and treatment group. Rosenbaum and Rubin (1983) introduced the propensity score, or the probability of being assigned to treatment given all relevant covariates, as a method to adjust for unbalanced covariates in an observational study. They proved that stratifying on the propensity score can lead to unbiased estimates of the treatment effect under strong ignorability, the assumption that potential outcomes and treatment assignment are independent given relevant group and individual level covariates

$$\Pr[Z_i = z_i | X_i, y_i(.)] = f_{Z_i|X_i}(z_i|X_i) \tag{1.1}$$

in which $Z_i$ denotes the assignment of treatments in group $i$, $y_i()$ the potential outcomes in group $i$, and $X_i$ the relevant covariates in group $i$.

Under strong ingnorability, Hong and Raudenbush (2006) use propensity scores to approximate a two-stage experiment using observational data. In the first stage, they use propensity scores to classify schools as low retention or high retention. In the second stage, the propensity score is used within high retention schools to separate students into seven strata based on their individual propensities to be retained.

### 1.5.3 Inverse Probability Weighting

Consider an observational study examining the impact of an exposure (yes/no) on an outcome (yes/no) that assumes no interference. Suppose we observe the data in Table 1.3.

In evaluating the effect of the exposure, our interest lies in the parameter $\tau = \sum_{j=1}^{n}\{y_j(0) - y_j(1)\}$, the sample average causal effect. As the outcome is binary, $\tau \in [-1, 1]$, where values close to 1 indicate significant preventive effects, values close to 0 indicate no effect, and values close to -1 indicate harm. Inference on $\tau$ can be accomplished using the following inverse

Table 1.3: Example exposure and outcome data for 8 individuals where $\tau = 0$. Observed data in the potential outcome structure are bolded.

| $y_j(0)$ | $y_j(1)$ | $Y_j$ | $Z_j$ | $X_j$ | $\hat{\Pr}[Z_j = z\|X_j]$ |
|---|---|---|---|---|---|
| **1** | 1 | 1 | 0 | 10 | 0.73 |
| **1** | 1 | 1 | 0 | 9 | 0.68 |
| **1** | 1 | 1 | 0 | 10 | 0.73 |
| **0** | 0 | 0 | 0 | 2 | 0.25 |
| 1 | **1** | 1 | 1 | 3 | 0.69 |
| 0 | **0** | 0 | 1 | 10 | 0.27 |
| 0 | **0** | 0 | 1 | 2 | 0.75 |
| 0 | **0** | 0 | 1 | 3 | 0.69 |

probability weighted (IPW) estimator:

$$\hat{\tau}^{IPW} = \frac{1}{n}\sum_{j=1}^{n}\frac{1\{Z_j = 0\}Y_j}{\Pr[Z_j = 0]} - \frac{1}{n}\sum_{j=1}^{n}\frac{1\{Z_j = 1\}Y_j}{\Pr[Z_j = 1]} \tag{1.2}$$

It is straightforward to show that $\hat{\tau}^{IPW}$ is unbiased, as

$$\begin{aligned} E[\hat{\tau}^{IPW}] &= \frac{1}{n}\sum_{j=1}^{n} E\left[\frac{1\{Z_j = 0\}Y_j}{\Pr[Z_j = 0]}\right] - \frac{1}{n}\sum_{j=1}^{n} E\left[\frac{1\{Z_j = 1\}Y_j}{\Pr[Z_j = 1]}\right] \\ &= \frac{1}{n}\sum_{j=1}^{n}\frac{\Pr[Z_j = 0]y_j(0)}{\Pr[Z_j = 0]} - \frac{1}{n}\sum_{j=1}^{n}\frac{\Pr[Z_j = 1]y_j(1)}{\Pr[Z_j = 1]} \\ &= \tau \end{aligned} \tag{1.3}$$

One of the key strong points of a randomized trial is that $\Pr[Z_j = z]$ is known in advance, so that for example if individuals are randomized with probability 1/2 to exposure, then $\hat{\tau}^{IPW} = \sum_{j=1}^{n}(1 - Z_j)Y_j/(n/2) - \sum_{j=1}^{n} Z_jY_j/(n/2)$, the familiar difference in proportions. In an observational study, $\Pr[Z_j = z]$ is rarely known, and must be estimated. If one is willing to invoke strong ignorability, i.e., the assumption in (1.1), then $\Pr[Z_j = z|X_j]$ can be used in place of $\Pr[Z_j = z]$ in $\hat{\tau}^{IPW}$. The intuition here is that $\Pr[Z_j = z|X_j]$ adjusts for situations where individuals with certain $X$ values are more likely to have $Z = z$, for example individuals closer to a town could be more likely to have access to a bed net. For individual $i = 1, ..., n$, $\Pr[Z_j = z|X_j]$ can be estimated using a statistical model, e.g., logistic regression.

The model predicted probabilities, $\hat{\Pr}[Z_j = z|X_j]$ are displayed in Table 1.3. Using

$\hat{\Pr}[Z_j = z|X_j]$ in the denominators in (4.4), $\hat{\tau}^{IPW} = 0.34$. If we had naively assumed $\Pr[Z_j = z|X_j] = 1/2$ for all $i$, $\hat{\tau}^{IPW} = 0.5$. The naive estimate is biased, but the IPW estimator offers an improvement toward the true $\tau = 0$.

## 1.6 Interference

In the motivating example, it is clear that the outcome of individual $j$ in group $i$ (disease) is dependent upon $z_i$ (vaccine status of other members of the bari) as well as $z_{ij}$ (individual vaccine status), i.e., interference is present. In addition to infectious diseases, interference is found in other settings. Hong and Raudenbush (2006) investigated the effect of kindergarten retention programs on math and reading test scores in those retained and those promoted to first grade. "Peer effects" (interference) are hypothesized to be present because it is plausible that a child promoted to first grade may flourish more in a learning environment in which low achieving potential classmates have been retained in kindergarten. Peer effects are also found in college roommates' grade point averages and decisions to join a fraternity (Sacerdote, 2001). Sobel (2006) examined neighborhood effects in the Moving to Opportunity study, a housing mobility experiment sponsored by the U.S. Department of Housing and Urban Development in which eligible ghetto residents in five U.S cities are randomly assigned to receive (or not) various forms of assistance to relocate. Interference is present because a household's decision to move or not may be influenced by whether or not their neighbors receive a housing voucher to move. Rosenbaum (2007) gives other detailed examples of experiments where interference is possible.

If interference is unrestricted, the set $\{y_{ij}(z) : z \in \mathcal{Z}\}$ may contain up to $2^n$ elements, rendering inference impracticable. A scientifically reasonable assumption in many settings is that interference can only occur between individuals in the same group (Halloran and Struchiner, 1991, 1995), i.e., the outcome $y_{ij}$ is only dependent on $z_i$, the $n_i$ dimensional vector of treatment assignments in group $i$. Sobel (2006) called this a partial interference assumption. Under this assumption, the potential outcomes of individual $j$ in group $i$ can be denoted by $\{y_{ij}(z) : z \in \mathcal{Z}_i\}$, where $\mathcal{Z}_i$ is the set of $2^{n_i}$ realizations of the randomly assigned

treatment vector in group $i$, $Z_i$.

Hudgens and Halloran (2008) defined estimands of interest in the presence of partial interference, specifically in the context of a two-stage randomized vaccine intervention to prevent disease. The concepts easily generalize to other two-stage randomized experiments. In the first stage, treatment strategies are randomized to groups, and in the second stage, individuals are randomized to treatment or no treatment based on the strategy of the group. An example could be that $\alpha_1$ is the strategy where half of a group is randomized to treatment and $\alpha_0$ is the strategy where one third of a group is randomized to treatment. The presence of two strategies introduces the idea of a direct and indirect effect of treatment. A direct effect of treatment is the effect of treatment within a given strategy. An indirect (spillover) effect is the effect of one strategy relative to another in the controls.

Average potential outcomes summarize the set of potential outcomes for each individual. Let the average potential outcome for individual $j$ on treatment $z = 0, 1$ in group $i$ under treatment strategy $\alpha_s$, $s = 0, 1$ be defined as

$$\bar{y}_{ij}(z;\alpha_s) \equiv \sum_{\omega \in \mathcal{R}^{n_i-1}} y_{ij}(z_{ij} = z, z_{i(j)} = \omega) \Pr_s(Z_{i(j)} = \omega | Z_{ij} = z) \tag{1.4}$$

where $z_{i(j)}$ represents the vector of the $n_i - 1$ treatment assigments for all individuals in group $i$ except for individual $j$, and $\mathcal{R}^{n_i-1}$ is the set of all vectors of length $n_i - 1$ composed of 1s and 0s. Let $\bar{y}_i(z;\alpha_s) \equiv \sum_{j=1}^{n_i} \bar{y}_{ij}(z;\alpha_s)/n_i$ be the average potential outcome in group $i$ for treatment assignment $z$ under treatment strategy $\alpha_s$. Let $\bar{y}(z;\alpha_s) = \sum_{i=1}^{k} \bar{y}_i(z;\alpha_s)/k$ be the population average potential outcome for treatment assignment $z$ under treatment strategy $\alpha_s$.

Additionally, define the average potential outcome for individual $j$ in group $i$ on treatment strategy $\alpha_s$ as

$$\bar{y}_{ij}(\alpha_s) \equiv \sum_{\omega \in \mathcal{R}^{n_i}} y_{ij}(z_i = \omega) \Pr_s(Z_i = \omega) \tag{1.5}$$

so that the average potential outcome in group $i$ under treatment strategy $\alpha_s$ is $\bar{y}_i(\alpha_s) \equiv \sum_{j=1}^{n_i} \bar{y}_{ij}(\alpha_s)/n_i$, and the population average potential outcome under treatment strategy $\alpha_s$

is $\bar{y}(\alpha_s) \equiv \sum_{i=1}^{k} \bar{y}_i(\alpha_s)/k$.

Given the definitions of the average potential outcomes above, Table 1.4 sums up the targets of inference at the individual, group, and population levels.

Table 1.4: Causal effects at the individual, group, and population level

| | Level | | |
|---|---|---|---|
| Causal Effect | Individual | Group | Population |
| Direct | $\bar{y}_{ij}(0;\alpha_s) - \bar{y}_{ij}(1;\alpha_s)$ | $\bar{y}_i(0;\alpha_s) - \bar{y}_i(1;\alpha_s)$ | $\bar{y}(0;\alpha_s) - \bar{y}(1;\alpha_s)$ |
| Indirect | $\bar{y}_{ij}(0;\alpha_0) - \bar{y}_{ij}(0;\alpha_1)$ | $\bar{y}_i(0;\alpha_0) - \bar{y}_i(0;\alpha_1)$ | $\bar{y}(0;\alpha_0) - \bar{y}(0;\alpha_1)$ |
| Total | $\bar{y}_{ij}(0;\alpha_0) - \bar{y}_{ij}(1;\alpha_1)$ | $\bar{y}_i(0;\alpha_0) - \bar{y}_i(1;\alpha_1)$ | $\bar{y}(0;\alpha_0) - \bar{y}(1;\alpha_1)$ |
| Overall | $\bar{y}_{ij}(\alpha_0) - \bar{y}_{ij}(\alpha_1)$ | $\bar{y}_i(\alpha_0) - \bar{y}_i(\alpha_1)$ | $\bar{y}(\alpha_0) - \bar{y}(\alpha_1)$ |

Hudgens and Halloran (2008) derived unbiased estimators for the group and population level effects in Table 1.4 in a completely randomized experiment. In a completely randomized experiment, exactly $m_i$ of $n_i$ individuals are randomized to treatment in the second stage, where $m_i$ is determined by the treatment strategy a group receives in the first stage. Let $\mathcal{R}_{m_i}^{n_i}$ denote the set of vectors $z_i$ composed of exactly $m_i$ 1s and $n_i - m_i$ 0s.

Hudgens and Halloran (2008) showed that estimators of the variances of the unbiased estimators do not exist without further assumptions. They derived variance estimators that are unbiased if causal effects are additive and are positively biased under the assumption of stratified interference. Stratified interference assumes that an individual's potential outcome on treatment $z \in \{0, 1\}$ is the same as long as a fixed number of other members in the group are assigned treatment, or that

$$y_{ij}(z_i) = y_{ij}(z_i') \text{ for all } z_i, z_i' \in \mathcal{R}_{m_i}^{n_i} \text{ such that } z_{ij} = z_{ij}' \tag{1.6}$$

The assumption in (3.5) is plausible in a wide range of settings. Liu and Hudgens (2014) identify conditions necessary to apply the parameter estimators and variance estimators of Hudgens and Halloran (2008) to form large-sample $1 - \alpha$ confidence intervals for the four population level effects in Table 1.4.

Other approaches have also been taken for statistical inference with interference. Rosen-

baum (2007) extends the covariance adjustment procedures to interference by comparing the observed data to what would have happened in a uniformity trial on the same group of subjects. A uniformity trial is an experiment in which all individuals receive placebo. The target of inference is the unobserved random variable $F$, the number of times that treated responses exceed control responses in the actual experiment but not the uniformity trial minus the number of times that treated responses exceed control responses in the uniformity trial but not in the actual trial. $F$ measures whether or not in an experiment with interference there is a greater tendency for treated subjects to have higher responses than controls than would have been seen in a uniformity trial with no effect. An upper $1-\alpha$ confidence interval for $F$ is found using the fact that $F$ can be decomposed into the difference of two Mann-Whitney statistics; one for the actual trial, $T$, and one for a uniformity trial, $\tilde{T}$. The key observation is that the distribution of $\tilde{T}$ is known because the null hypothesis of no effect is true in the uniformity trial.

This method is extended to incorporate baseline covariate information following similar logic to Rosenbaum (2002a). Let $\tilde{\epsilon}(Y + Z\tau^0) = \tilde{\epsilon}(y(0)) = e$ be the residuals from some model of $y(0)$ as a function of $X$. The residuals $e$ are used in the computation of the ranks in the Mann-Whitney statistic from the actual trial, $T_e$. An upper $1-\alpha$ confidence interval for $F$ follows in the same manner as in the previous paragraph.

For a two-stage randomized trial that assumes partial interference, Tchetgen Tchetgen and VanderWeele (2012) derived exact confidence intervals for the four effects of interest. In an observational study, they use inverse probability weighted (IPW) estimators to derive unbiased estimators of the four population effects in Table 1.4. Specifically, assuming that for $s = 0, 1$, $\pi_i(z_i; \alpha_s) = \prod_{j=1}^{n_i} \alpha_s^{z_{ij}}(1-\alpha_s)^{1-z_{ij}}$, for $z = 0, 1$ is the mechanism by which individuals in group $i$ are randomized to treatment, let

$$\hat{Y}_i^{ipw}(z; \alpha_s) = \frac{\sum_{j=1}^{n_i} \pi_i(Z_{i(j)}; \alpha_s) 1\{Z_{ij} = z\} Y_{ij}(Z_i)}{n_i \times f_{Z_i|X_i}(z_i|X_i)} \tag{1.7}$$

and

$$\hat{Y}_i^{ipw}(\alpha_s) = \frac{\sum_{j=1}^{n_i} \pi_i(Z_i;\alpha_s) Y_{ij}(Z_i)}{n_i \times f_{Z_i|X_i}(z_i|X_i)} \tag{1.8}$$

Assuming strong ignorability (1.1), and that $\Pr[Z_i = z_i|X_i] > 0$ for all $z_i \in \mathcal{Z}_i$, they show that $\hat{Y}_i^{ipw}(z;\alpha_s)$ is an unbiased estimator of $y_i(z;\alpha_s)$ and that $\hat{Y}_i^{ipw}(\alpha_s)$ is an unbiased estimator of $y_i(\alpha_s)$, and thus that $k^{-1}\sum_{i=1}^{k}\{\hat{Y}_i^{ipw}(0;\alpha_s) - \hat{Y}_i^{ipw}(1;\alpha_s)\}$ is an unbiased estimator of the population direct effect of treatment strategy $\alpha_s$, $k^{-1}\sum_{i=1}^{k}\{\hat{Y}_i^{ipw}(0;\alpha_0) - \hat{Y}_i^{ipw}(0;\alpha_1)\}$ is an unbiased estimator of the population indirect effect, $k^{-1}\sum_{i=1}^{k}\{\hat{Y}_i^{ipw}(0;\alpha_0) - \hat{Y}_i^{ipw}(1;\alpha_1)\}$ is an unbiased estimator of the population total effect, and $k^{-1}\sum_{i=1}^{k}\{\hat{Y}_i^{ipw}(\alpha_0) - \hat{Y}_i^{ipw}(\alpha_1)\}$ is an unbiased estimator of the population overall effect. In observational studies, $f_{Z_i|X_i}(z_i|X_i)$ is rarely known, and must be estimated, e.g., by mixed effects logistic regression. Perez-Heydrich et al. (2014) study the asymptotic properties of the estimators in (1.7) and (1.8) when the propensity score is estimated, and apply them to the aforementioned cholera data in Bangladesh.

Bowers et al. (2013) use a parametric model to reveal the unobserved potential outcomes. They encourage researchers to write down their own models and they provide an inferential algorithm. Model parameters $\theta$ are chosen to have meaning in the context of interference. Each value of $\theta$ generates a hypothesis that can be tested by a p-value found using principles of randomization based inference. A $1-\alpha$ confidence interval for $\theta$ is formed by those values of $\theta$ with p-values greater than or equal to $\alpha$.

For illustration, let $z$ denote any possible $n \times 1$ vector realization of the random treatment assignment vector $Z$ for $z \in \mathcal{Z}$, where $\mathcal{Z}$ denotes the set of all possible treatment assignment vectors. In each vector $z$, $z_i = 1$ if individual $i$ receives treatment and $z_i = 0$ if individual $i$ receives control for $i = 1, ..., n$. Let $Z = 0$ denote the vector where all individuals are assigned control, or a uniformity trial in the language of Rosenbaum (2007). Let $y(z)$ denote the $n \times 1$ vector of potential outcomes for randomization $z$, and let $y_i(z)$ denote the $i^{th}$ element of $y(z)$ for $i = 1, ..., n$. Now, posit a model that transforms potential outcomes in a uniformity trial

to the potential outcomes revealed by any randomization $z \in \mathcal{Z}$:

$$y_i(z) = \{\beta + (1 - z_i)(1 - \beta)\exp(-\tau^2 z' S_i)\} y_i(0) \tag{1.9}$$

in which $S_i$ is the $i^{th}$ row of $S$, an $n \times n$ adjacency matrix where $S_{ij} = 1$ if units $i$ and $j$ are linked and 0 otherwise, and $\theta = (\beta, \tau)$. In (1.9), $\{\beta + (1 - z_i)(1 - \beta)\exp(-\tau^2 z' S_i)\}$ can be thought of as the multiplicative effect of treatment program $z$ on subject $i$'s potential outcome in a uniformity trial in the presence of interference. If $z_i = 1$, this multiplicative effect simplifies to $\beta$, the direct effect of treatment. If $z_i = 0$, this multiplicative effect is a function of $\beta$ and $\tau$, a parameter measuring the spillover effect. If $\tau = 0$, then individuals with $z_i = 0$ have a multiplicative effect of 1, or do not receive any benefit from treatment. The quantity $z' S_i$ is the number of connections individual $i$ has to other members in the group.

The experiment assigns $Z = \tilde{z}$, and thus reveals the potential outcome vector $y(\tilde{z})$. Assuming (1.9), $y_i(0) = \{\beta + (1 - \tilde{z}_i)(1 - \beta)\exp(-\tau^2 \tilde{z}' S_i)\}^{-1} y_i(\tilde{z})$ for $i = 1, ..., n$ and thus we can generate any $y_i(z)$ for $z \in \mathcal{Z}$ as

$$y_i(z) = \{\beta + (1 - z_i)(1 - \beta)\exp(-\tau^2 z' S_i)\}\{\beta + (1 - \tilde{z}_i)(1 - \beta)\exp(-\tau^2 \tilde{z}' S_i)\}^{-1} y_i(\tilde{z}) \tag{1.10}$$

for $i = 1, ..., n$ for any $z \in \mathcal{Z}$. Hypotheses about $\beta$ and $\tau$ can be tested using randomization based inference as follows. Under any null hypothesis, all potential outcomes are generated by (1.10). A test statistics $T$ is chosen such that its absolute value is large when treated and control outcome distributions are different and small when treated and control outcome distributions are similar. Examples of choices for $T$ include the difference in means, differences in sums of ranks, or the Kolmogorov-Smirnov test statistic. To measure the extremeness of the observed data under this null hypothesis, compute the p-value as

$$p_{(\beta,\tau)} = \sum_{z \in \mathcal{Z}} 1\{|T(z, y(z))| \geq |T^{obs}|\} / |\mathcal{Z}| \tag{1.11}$$

where $|\mathcal{Z}|$ is the number of elements in $\mathcal{Z}$ and $T^{obs} = T(\tilde{z}, y(\tilde{z}))$. The hypothesis is rejected if $p_{(\beta,\tau)}$ is less than $\alpha$. Following the standard procedure of inverting a test, $1 - \alpha$ confidence

sets for $\beta$ and $\tau$ are composed of all values of $\beta$ and $\tau$ for which $p_{(\beta,\tau)}$ is greater than or equal to $\alpha$.

Toulis and Kao (2013) estimate causal peer influence effects using a frequentist and a Bayesian procedure. In their language, individual $i = 1, ..., N$ is a member (node) of the vertex set in the network. For each individual $i$, there are $n_i = |\mathcal{N}_i|$ neighboring nodes who can influence $Y_i$, the outcome of interest, where $\mathcal{N}_i$ is the set of all neighboring nodes. Let $Z_i$ denote the treatment assignment of individual $i$, and let the vector $Z$ denote the treatment assignments of all individuals in the network. The potential outcome of individual $i$ under treatment $Z$ is $Y_i(Z) \equiv Y_i(Z_i, Z_{(i)})$, where $Z_{(i)}$ is the vector of treatment assignments excluding individual $i$. It is clear from the structure of the network that $Y_i(Z)$ is equal to $Y_i(Z_i, Z_{\mathcal{N}_i})$, where $Z_{\mathcal{N}_i}$ are the treatment assignments for the neighboring nodes. Let $Y_i(0) \equiv Y_i(Z_i = 0, Z_{\mathcal{N}_i} = 0)$. The causal estimand of primary effects is

$$\eta \equiv N^{-1} \sum_{i=1}^{N} \{Y_i(1, Z_{\mathcal{N}_i} = 0) - Y_i(0)\} \tag{1.12}$$

The causal estimand of $k$-level peer effects is

$$\delta_k \equiv |V_k|^{-1} \sum_{i \in V_k} \left[ \binom{n_i}{k}^{-1} \sum_{z \in Z(\mathcal{N}_i, k)} Y_i(0, z) - Y_i(0) \right] \tag{1.13}$$

where $V_k$ is the set of all nodes that have at least $k$ neighbors, or the set $\{i : n_i \geq k\}$, and $Z(\mathcal{N}_i, k)$ is the set of all assignments for nodes $\mathcal{N}_i$ in which exactly $k$ receive treatment. Additionally, they define estimands of "insulated neighbors" and "non-insulated neighbors" which involve shared neighbors between nodes. In the Bayesian linear model approach for inference on (1.12) and (1.13), consider a linear model for the individual potential outcomes:

$$Y_i(Z) = \mu + \tau Z_i + \gamma a_i' Z + \epsilon_i \tag{1.14}$$

where $a_i$ is the $i^{th}$ column vector of the adjacency matrix $A$ that summarizes links between units such that $A_{ij} = 1$ if units $i$ and $j$ are linked and 0 otherwise, and it is assumed that the

$N \times 1$ vector $\epsilon \sim \mathcal{N}(0, \sigma^2 I)$ is independently and identically distributed noise. Let $S = A'Z$ be the $N \times 1$ vector that summarizes the amount of exposure to peer influence for each node. Under (1.14), the causal estimand of primary effects is

$$\begin{aligned} \eta &= N^{-1} \sum_i [Y_i(1, z = 0) - Y_i(0)] \\ &= N^{-1} \sum_i \tau = \tau \end{aligned} \tag{1.15}$$

and the causal estimand of $k$-level peer effects is

$$\begin{aligned} \delta_k &= |V_k|^{-1} \sum_{i \in V_k} \left[ \binom{n_i}{k}^{-1} \sum_{z \in Z(\mathcal{N}_i, k)} Y_i(0, z) - Y_i(0) \right] \\ &= |V_k|^{-1} \sum_{i \in V_k} \left[ \binom{n_i}{k}^{-1} \sum_{z \in Z(\mathcal{N}_i, k)} S_i(z)\gamma \right] \\ &= k\gamma |V_k|^{-1} \sum_{i \in V_k} W_i \propto \gamma \end{aligned} \tag{1.16}$$

The authors describe a Bayesian hierarchical model for inference on $\tau$ and $\gamma$ which in turn leads to inference on $\eta$ and $\delta_k$.

## 1.7 Bayesian Causal Inference

Rubin (1978) investigated the role of randomization in Bayesian inference for causal effects. In this view, potential outcomes are no longer fixed, but are considered random variables. Inference on causal parameters will be accomplished by modeling the missing potential outcomes as a function of covariates, revealed potential outcomes, and parameters. Inference on causal parameters is carried out by posterior distributions rather than confidence intervals. There are no hypothesis tests to invert.

For illustration, consider $Y(0)$, and $Y(1)$, now random variables, as the potential outcomes of interest, and let $X$ be the covariates and $\theta$ the parameters. Following Rubin (1978), Bayesian causal inference proceeds by specifying a joint distribution of all observ-

able quantities, $\Pr[Y(0), Y(1), Z, X|\theta]$, reasonable priors for $\theta$, and treating the analysis as a missing data problem. The missing values of interest are the missing potential outcomes, $Y^{mis} = Z'Y(0) + (1 - Z)'Y(1)$. It is typical that the joint distribution is factored into two important distributions:

$$\Pr[Y(0), Y(1), Z, X|\theta] = \Pr[Y(0), Y(1), X|\theta] \Pr[Z|Y(0), Y(1), X, \theta] \tag{1.17}$$

where $\Pr[Y(0), Y(1), X|\theta]$ is a model of "the science" and $\Pr[Z|Y(0), Y(1), X, \theta]$ is the assignment mechanism. If the assignment mechanism is unconfounded and ignorable as in (1.1), then $\Pr[Z|Y(0), Y(1), X] = \Pr[Z|X]$. This assumption is common in many observational studies. After specifying the necessary inputs, samples are taken from the posterior distribution of the causal effect using a Gibbs sampler.

1. Sample initial guesses for $\theta$ from the prior $f(\theta)$.

2. Sample values for $Y^{mis}$ conditional on $\theta$ from $f(Y^{mis}|Y^{obs}, X^{obs}, Z, \theta)$. Causal effects that are functions of $Y$, e.g., $n^{-1} \sum_i \{Y_i(0) - Y_i(1)\}$, can be recorded.

3. Sample values for $\theta$ from $f(\theta|Y^{obs}, Y^{mis}, X^{obs}, Z)$, i.e., conditional on the sampled $Y^{mis}$ in step 2. Causal effects that are functions of $\theta$, e.g., $E[Y(0)] - E[Y(1)]$, can be recorded.

Bayesian inference for causal effects is flexible in the sense that the finite or superpopulation persective can be adopted.

Applications of Bayesian causal inference abound. Chib and Hamilton (2000) used Bayesian causal inference for clustered data. Dawid (2000) uses Bayesian decision analysis for causal inference. Schwartz et al. (2011) put forth a Bayesian approach for causal inference on intermediate variables. Crowley et al. (2014) perform a Bayesian causal analysis on the effects of Haloperidol in paired mice where SUTVA is an untenable assumption.

To our knowledge, the literature on Bayesian methods for causal inference in the presence of interference is limited. Toulis and Kao (2013) estimate causal peer influence effects using a frequentist and a Bayesian procedure.

## 1.8 Summary of Research

In the first paper, two exact confidence intervals are derived for the group and population direct effects under no interference when the outcome is binary, as in a vaccine trial. Rosenbaum (2001) provided a key ingredient, attributable effects, to be used in the creation of one of the exact confidence intervals. The other exact confidence interval makes use of the permutation test strategy outlined above for Table 1.2.

In the second paper, exact confidence intervals are developed for the population direct, indirect, total, and overall effects in a two stage randomized trial that assumes stratified interference. These methods will be contrasted with the exact confidence intervals for the four effects of Tchetgen Tchetgen and VanderWeele (2012) in the presence of stratified interference.

In the third paper, a Bayesian approach is proposed for inference on causal parameters in observational data where interference is present. The motivating data arise from the 2007 Demographic and Health Survey in the Democratic Republic of the Congo, an observational study. In the Messina et al. (2011) analysis, 7746 individuals in 300 communities with complete individual and community covariate and outcome data were included in a multilevel statistical model. The group level parameter for bed net use on malaria outcome was statistically significant, whereas the analagous individual level parameter was not. In a re-analysis of these data, Bayesian methods are developed for inference on the population direct, indirect, total, and overall effect of bed net use on malaria outcome.

# CHAPTER 2: RANDOMIZATION INFERENCE FOR TREATMENT EFFECTS ON A BINARY OUTCOME

## 2.1 Introduction

In many settings inference is desired about the effect of a treatment relative to the absence of treatment on a particular outcome. In studies where treatment is randomly assigned, randomization based inference can be employed to draw conclusions about the effect of treatment. For instance, when the outcome is continuous, randomization based confidence intervals can be formed using the classic approach of Hodges and Lehmann (1963). In addition to randomization, this approach relies on one particular key assumption, namely that the effect of treatment is additive, i.e., the same for all individuals. Additivity is a strong assumption that may not hold in many settings, particularly if the outcome is binary (LaVange et al., 2005). In this paper, two methods are developed for constructing randomization based confidence sets for the average effect of treatment on a binary outcome without assuming additivity. These sets are formed by *(i)* combining prediction sets for attributable effects (Rosenbaum, 2001), and by *(ii)* inverting a permutation test.

Specifically, consider a study in which $m$ of $n$ individuals are randomized to treatment and subsequently a binary outcome is measured. Let the binary outcome of interest be denoted by $Y_j$ where $Y_j = 1$ if the event occurs and 0 otherwise for individuals $j = 1, \ldots, n$. Let treatment assignment be indicated by $Z_j$ where $Z_j = 1$ if treatment and 0 if placebo. Prior to treatment assignment, assume each individual has two potential outcomes: $y_j(1)$ if assigned treatment, and $y_j(0)$ if placebo (or control). After treatment assignment, one of the two potential outcomes is observed so that the observed outcome for individual $j$ is $Y_j = Z_j y_j(1) + (1 - Z_j) y_j(0)$. Let $Z$ denote the vector of treatment assignments, $Y$ denote the vector of observed outcomes, and $y(z)$ denote the vector of potential outcomes when

all $n$ individuals are assigned $z \in \{0,1\}$. Define the treatment effect for individual $j$ to be $\delta_j = y_j(1) - y_j(0)$, so that $\delta_j = 1$ if treatment causes event, 0 if treatment has no effect, and $-1$ if treatment prevents event. Let $\delta = y(1) - y(0)$ be the vector of treatment effects, and let $\tau = \sum \delta_j / n$ be the average treatment effect, where here and in the sequel $\sum = \sum_{j=1}^{n}$. Our goal is to construct a confidence set for $\tau$.

In both of the methods to follow, inference on $\delta$ will be used as a starting point for inference on $\tau$. Prior to seeing the data, $\delta \in \{-1,0,1\}^n$, a set with $3^n$ elements. Once the data are observed, one of the two potential outcomes is revealed and one is missing. Because the missing outcome is known to equal 0 or 1, once the data are observed $\delta_j$ is restricted to take one of two values for each individual $j$, such that there are only $2^n$ $\delta$ vectors compatible with the observed data. Similarly, prior to observing the data, the parameter $\tau$ can take on values in $\{-n/n, \ldots, 0/n, \ldots, n/n\}$, a set with $2n+1$ elements of width two, where here and in the sequel we define the width of a set to be the difference between the maximum and minimum values of the set. After observing the data, it can be easily shown that the set of compatible $\tau$ values is

$$\left\{ \frac{\sum Y_j(2Z_j - 1) - m}{n}, \frac{\sum Y_j(2Z_j - 1) - m + 1}{n}, \ldots, \frac{\sum Y_j(2Z_j - 1) - m + n}{n} \right\} \tag{2.1}$$

a set with $n+1$ elements of width one. Each of the $2^n$ compatible $\delta$ vectors maps to one of these $n+1$ compatible $\tau$ values. The data are informative in the sense that $n$ of the possible $\tau$ values can be rejected (with type I error zero). On the other hand, the null $\tau$ value of 0 will always be contained in the set of compatible $\tau$ values. This is analogous to a well known result about "no assumption" large sample treatment effect bounds (Manski, 1990). The methods below construct confidence sets for $\tau$ that are subsets of the set (4.9) and thus potentially of width less than one.

The two proposed methods are similar in spirit to the classic Hodges-Lehmann confidence interval in that randomization-based tests are inverted to construct the confidence sets. However, unlike the Hodges-Lehmann approach, no assumption is made that the effect is additive. This is critical because in many settings it will be unlikely or implausible that the treatment

effect is the same for all individuals. For example, to assume $\delta_j = 1$ for all $j$ corresponds to the scenario $y_j(1) = 1$ and $y_j(0) = 0$ for all $j$, i.e., everyone has an event if and only if treated. Moreover, this particular additivity assumption could be rejected with type I error zero if $Y_j = 0$ for at least one individual assigned treatment or $Y_j = 1$ for at least one individual assigned placebo. An analogous statement applies to the assumption that $\delta_j = -1$ for all $j$.

The two proposed methods rely on the randomization-based mode of inference wherein the $n$ individuals are viewed as the finite population of interest and probability arises only through the randomization assignment to treatment or placebo (Rosenbaum, 2002b, chap. 2). The randomization-based approach to inference has several appealing properties. For example, the resulting inferences are exact without relying on distributional assumptions and do not require large sample approximations. Randomization-based inference also does not require the observed data constitute a random sample from some infinite population, unlike the more common superpopulation model (Robins, 1988). This is important in settings where assuming random sampling from the target population may be dubious. For example, individuals who volunteer to participate in a clinical trial may be a biased sample from the general population. Similarly, animals or organisms in a laboratory experiment may differ fundamentally from their counterparts in nature. See Rosenbaum (2002b); Robins (1988); Miettinen and Cook (1981); Rubin (1991); Lehmann (1998) for additional discussion related to the various modes of inference for treatment (i.e., causal) effects.

The outline of the rest of this paper is as follows. In Section 2, an approach for finding a confidence set for $\tau$ based on attributable effects (Rosenbaum, 2001) is proposed. In Section 3, a confidence set for $\tau$ is found by inverting a permutation test. In Section 4 the two proposed confidence sets are compared with a large sample confidence interval for $\tau$ (Robins, 1988) as well as the usual Wald confidence interval and a commonly used exact interval for the difference in binomial proportions; the different confidence intervals (or sets) are evaluated in simulation studies and illustrated using data from a vaccine adherence trial. In Section 5, extensions to settings with more than one group are considered. Section 6 concludes with a

discussion.

## 2.2 Attributable Effect Sets

This section describes how a $1-\alpha$ confidence set for $\tau$ can be constructed by combining prediction sets for attributable effects (Rosenbaum, 2001). The observed data $\{Z, Y\}$ can be displayed in traditional $2 \times 2$ form as in Table 2.1. Noting that $\sum Z_j Y_j = \sum Z_j y_j(1)$, $\sum Z_j(1-Y_j) = \sum Z_j(1-y_j(1))$, $\sum(1-Z_j)Y_j = \sum(1-Z_j)y_j(0)$, $\sum(1-Z_j)(1-Y_j) = \sum(1-Z_j)(1-y_j(0))$, and $y_j(1) = y_j(0) + \delta_j$, Table 2.1 can be re-expressed as a function of $Z$, $y(0)$, and $A^1(Z,\delta) = \sum Z_j \delta_j$, the attributable effect of treatment in the treated (Rosenbaum, 2001), as shown in Table 2.2. In words, $A^1(Z,\delta) = \sum Z_j y_j(1) - \sum Z_j y_j(0)$ is the difference in the number of events which occurred in the treated subjects and the number of events that would have occurred if, contrary to fact, they had been exposed to control instead. After observing the data, it can be inferred that $A^1(Z,\delta) \in \{\sum Z_j Y_j - m, \sum Z_j Y_j - m + 1, \ldots, \sum Z_j Y_j\}$, a set with $m+1$ elements. The observed data can be used to construct a prediction set for $A^1(Z,\delta)$. We refer to these sets as prediction sets rather than confidence sets because $A^1(Z,\delta)$ is a random variable rather than a parameter. Rosenbaum (2001) described how to construct such prediction sets. In particular, consider testing $H_0 : \delta = \delta^0$ for some compatible vector of effects $\delta^0$. Under $H_0$, subtracting $A^1(Z,\delta^0)$ from the (1,1) cell of Table 2.2 and adding $A^1(Z,\delta^0)$ to the (1,2) cell creates a table with fixed margins, as the row margins of this "adjusted" table are fixed by design and the column margins are fixed because $\sum y_j(0)$ does not depend on $Z$. Let $U = \sum Z_j Y_j - A^1(Z,\delta) = \sum Z_j y_j(0)$ denote the number of events in the treated individuals had, contrary to fact, they not been treated. Note $U$ is pivotal because its distribution under $H_0$ does not involve $\delta^0$, i.e., $U$ follows a hypergeometric distribution with $\Pr(U=u) = \binom{\sum y_j(0)}{u}\binom{n-\sum y_j(0)}{m-u}\Big/\binom{n}{m}$ for $u \in \{\max\{0, m+\sum y_j(0) - n\}, \ldots, \min\{\sum y_j(0), m\}\}$. Let $u(\delta^0) = \sum Z_j Y_j - A^1(Z,\delta^0)$, the value of $U$ under $H_0$, and let the two-sided Fisher's exact test p-value be $p_{\delta^0}(Z,Y) = \sum_u \Pr(U=u) 1\{\Pr(U=u) \leq \Pr(U=u(\delta^0))\}$. Note each of the $2^n$ compatible $\delta^0$ corresponds to one of the $m+1$ compatible $A^1(Z,\delta^0)$. Therefore, those $\delta^0$ that map to the same value of $A^1(Z,\delta^0)$ will all yield the same p-value when testing $H_0$.

Let $\mathcal{P}(A^1(Z,\delta)) = \{A^1(Z,\delta) : p_\delta(Z,Y) \geq \alpha\}$ denote the set of compatible attributable effects of treatment in the treated where the null $H_0 : \delta = \delta^0$ is not rejected at significance level $\alpha$. The set $\mathcal{P}(A^1(Z,\delta))$ is a $1-\alpha$ prediction set for $A^1(Z,\delta)$ in the sense that $\Pr[A^1(Z,\delta) \in \mathcal{P}(A^1(Z,\delta))] \geq 1-\alpha$.

Similarly, define the attributable effect of treatment in the untreated as $A^0(Z,\delta) = \sum(1-Z_j)\delta_j$. In words, $A^0(Z,\delta) = \sum(1-Z_j)y_j(1) - \sum(1-Z_j)y_j(0)$ is the difference in the number of events in the control subjects had, contrary to fact, they been treated and the number of events actually observed in the control subjects. After observing the data, it can be inferred that $A^0(Z,\delta) \in \{-\sum(1-Z_j)Y_j, -\sum(1-Z_j)Y_j + 1, \ldots, -\sum(1-Z_j)Y_j + n - m\}$, a set with $n-m+1$ elements. A $1-\alpha$ prediction set can be constructed for $A^0(Z,\delta)$ in the same fashion as for $A^1(Z,\delta)$. While the attributable effects $A^1(Z,\delta)$ and $A^0(Z,\delta)$ are random variables, they are constrained in sum to equal a constant:

$$A^1(Z,\delta) + A^0(Z,\delta) = \sum Z_j\delta_j + \sum(1-Z_j)\delta_j = \sum \delta_j = n\tau \tag{2.2}$$

The relationship between the attributable effects and $\tau$ in (2.2) suggests combining prediction sets for $A^1(Z,\delta)$ and $A^0(Z,\delta)$ to obtain a confidence set for $\tau$. The following proposition indicates that a confidence set for $\tau$ can be formed by combining prediction sets with a Bonferroni type adjustment.

**Proposition 2.1.** *If $\{L^1, L^1+1, \ldots, U^1\}$ is a $1-\alpha/2$ prediction set for $A^1(Z,\delta)$, where $L^1$ is the minimum of the prediction set and $U^1$ is the maximum, and $\{L^0, L^0+1, \ldots, U^0\}$ is a $1-\alpha/2$ prediction set for $A^0(Z,\delta)$, where $L^0$ and $U^0$ are defined similarly, then $\{(L^1+L^0)/n, (L^1+L^0+1)/n, \ldots, (U^1+U^0)/n\}$ is a $1-\alpha$ confidence set for $\tau$.*

A proof of Proposition 2.1 is given in the Appendix. Constructing a confidence set for $\tau$ as described in Proposition 2.1 only requires testing $n+2$ hypotheses, as there are $m+1$ compatible values of $A^1(Z,\delta^0)$ that must be tested and there are $n-m+1$ compatible values of $A^0(Z,\delta_0)$ that must be tested. Thus the attributable effect based confidence set for $\tau$ is computationally feasible even for large $n$; this is in contrast to the permutation test approach

described next.

Note Proposition 2.1 relies on a Bonferroni type adjustment. Because $A^1(Z, \delta)$ and $A^0(Z, \delta)$ are constrained according to (2.2), it might be tempting to instead add the lower and upper bounds of two $1 - \alpha$ prediction sets and divide by $n$ (i.e., without a Bonferroni type adjustment). However, such a naive approach is not guaranteed to provide coverage of at least $1-\alpha$ as demonstrated by the following example. Suppose an experiment is to be conducted with $m = 4$ of $n = 9$ individuals to be assigned treatment. As each individual's pair of outcomes $\{y_j(0), y_j(1)\}$ can take on 4 values, there are $4^9$ possible sets of potential outcomes for the finite population of individuals. Each of these sets maps to one of the $2n + 1 = 19$ values of $\tau$. Consider the subset of these $4^9$ sets that map to $\tau = 1/9$. For each of the sets of potential outcomes in this subset, there are $\binom{n}{m}$ possible observed data sets. Applying the naive approach described above of combining two 95% prediction sets without a Bonferroni adjustment to each of the possible observed data sets, only 92% of the sets contain $\tau = 1/9$.

## 2.3 Inverted Permutation Test

A permutation based approach can also be employed to find a confidence set for $\tau$. Prior to specifying a null hypothesis $H_0 : \delta = \delta^0$, each individual has one observed and one missing potential outcome; however, under $H_0$, both outcomes are known. A null hypothesis with this property is considered sharp. If the missing outcome for individual $j$ is $y_j(0)$, it is known under the null to equal $y_j(1) - \delta_j^0 = Y_j - \delta_j^0$, and if the missing outcome is $y_j(1)$, it is known under the null to equal $y_j(0) + \delta_j^0 = Y_j + \delta_j^0$. To determine how likely the observed data are under $H_0$, a test statistic can be chosen, its distribution under the null computed, and a measure of extremeness of the observed data defined (Rubin, 1991§4.1). A natural choice for the test statistic is the difference in observed means

$$T = \sum Z_j Y_j / m - \sum (1 - Z_j) Y_j / (n - m) \tag{2.3}$$

Neyman (1923) showed that $T$ is an unbiased estimator of $\tau$, i.e., $E(T) = \tau$, where the expected value is taken over all possible hypothetical randomizations of $m$ of the $n$ individuals to treatment under the true $\delta$ vector. The sampling distribution of $T$ under the null can be determined exactly by computing $T$ for each of the $C = \binom{n}{m}$ possible randomizations because all potential outcomes are known under the sharp null $H_0$. For randomization $c = 1, \ldots, C$, let $t^c$ denote the value of $T$ under $H_0$. Each randomization occurs with probability $1/C$, so the permutation test p-value is defined to be $\sum_{c=1}^{C} 1\{|t^c - \tau^0| \geq |t^{obs} - \tau^0|\}/C$ where $t^{obs}$ is the value of $T$ for the observed data, and $\tau^0 = \sum \delta_j^0/n$. The subset of compatible $\delta^0$ vectors where the permutation test p-value is greater than or equal to $\alpha$ forms a $1 - \alpha$ confidence set for $\delta$. The $\tau^0$ values corresponding to the $\delta^0$ vectors in this confidence set for $\delta$ form a $1 - \alpha$ confidence set for $\tau$.

Although finding a confidence set for $\delta$ entails explicitly testing $2^n$ hypotheses, finding a confidence set for $\tau$ can be accomplished by testing only $O(n^4)$ hypotheses. To see this, let $n_{zy} = \sum_{j=1}^{n} 1\{Z_j = z, Y_j = y\}$ for $z \in \{0, 1\}$ and $y \in \{0, 1\}$. For the $n_{11}$ individuals with $Z_j = 1$ and $Y_j = 1$, $\delta_j$ can be 0 or 1. Holding the $\delta_j$ value fixed for the other $n_{10} + n_{01} + n_{00}$ individuals, for fixed $v \in \{0, 1, \ldots, n_{11}\}$ all $\delta$ vectors with $\sum_{j:Z_j=Y_j=1} 1\{\delta_j = 1\} = v$ will lead to the same $\tau$ value and permutation p-value, i.e., it is sufficient to test $n_{11} + 1$ hypotheses about individuals with $Z_j = Y_j = 1$. Similar logic can be applied to the other three cross-classifications of treatment and outcome, such that it is sufficient to test $(n_{11} + 1)(n_{10} + 1)(n_{01} + 1)(n_{00} + 1)$ hypotheses to find a confidence set for $\tau$.

As $O(n^4)$ becomes large, computing permutation confidence sets may become infeasible. In addition to utilizing the `compiler` package (R Core Team, 2014), the following two strategies may be employed to improve computational efficiency. First, rather than using all $\binom{n}{m}$ possible randomizations to find the permutation p-value for each hypothesis being tested, a Monte Carlo procedure based on a random sample of the randomizations can be employed to approximate the p-value (Mehta and Patel, 2003). Second, the lower limit of the confidence set for $\tau$ can be found as follows. Starting with the smallest compatible $\tau$ value, compute the permutation p-value for each corresponding $\delta$ vector. If at least one p-value is greater

than or equal to $\alpha$, set the lower limit to this value of $\tau$. Otherwise, repeat this process for the next largest compatible $\tau$ value until a corresponding $\delta$ vector is found whose p-value is greater than or equal to $\alpha$. The upper limit can be found analogously starting with the largest compatible $\tau$ value.

## 2.4 Illustrations

### 2.4.1 Simple Examples

In this section, the attributable effects and permutation confidence sets for $\tau$ are compared with an asymptotic confidence interval for $\tau$. Robins (1988) proposed the following large sample $(1-\alpha)$ confidence interval for $\tau$

$$T \pm z_{(1-\alpha/2)}\{\hat{p}_1(1-\hat{p}_1)/m + \hat{p}_0(1-\hat{p}_0)/(n-m) + \hat{R}\}^{1/2} \tag{2.4}$$

where $\hat{p}_1 = \sum Z_j Y_j/m$, $\hat{p}_0 = \sum(1-Z_j)Y_j/(n-m)$, $\hat{R} = \{(2\hat{p}_0 - \hat{p}_1)(1-\hat{p}_1) - \hat{p}_0(1-\hat{p}_0)\}/n$ if $\hat{p}_1 \geq \hat{p}_0$, $\hat{R} = \{(2\hat{p}_1 - \hat{p}_0)(1-\hat{p}_0) - \hat{p}_1(1-\hat{p}_1)\}/n$ if $\hat{p}_0 > \hat{p}_1$, and $z_{(1-\alpha/2)}$ denotes the $1-\alpha/2$ quantile of a standard normal distribution. As $n \to \infty$ with $m/n \to c \in (0,1)$, the interval (2.4) will contain $\tau$ with probability $1-\alpha$ (Robins, 1988).

To compare the methods, consider an experiment with $m = 4$ of $n = 8$ individuals assigned treatment. As each individual's outcomes $\{y_j(0), y_j(1)\}$ can take on 4 values, there are $4^8$ possible sets of potential outcomes for the finite population of individuals. For each of these $4^8$ sets, there are $\binom{n}{m}$ possible observed data sets. For each of the $4^8\binom{n}{m}$ possible combinations of potential outcomes and observed data sets, attributable effects and permutation confidence sets and asymptotic confidence intervals were computed. Figure 2.1 displays the coverage probability and average width for the three methods at each of the $2n+1 = 17$ values of $\tau$ for $\alpha = 0.05$. To illustrate how the points in Figure 2.1 were computed, consider the coverage probability of the asymptotic confidence set for $\tau = -6/8$ in the top panel of Figure 2.1. Of the $4^8 = 65536$ sets of potential outcomes, 120 have $\tau = -6/8$. For these 120 sets of potential outcomes, the asymptotic sets has coverage probability 0.79 for 28 of the sets, 0.71

for 64 of the sets, and 0.79 for 28 of the sets, so the coverage probability for the asymptotic confidence set at $\tau = -6/8$ is the weighted mean, 0.75. The asymptotic confidence sets fail to provide the desired 95% coverage for many $\tau$ values; on the other hand, the attributable effects and permutation confidence sets provide the desired level of coverage for all $\tau$ values. Permutation confidence sets have a smaller width than the attributable effects confidence sets for each value of $\tau$ in this experiment.

### 2.4.2 Simulation Study

To further study the proposed methods, the permutation, attributable effects, and asymptotic approaches were compared to the usual Wald interval

$$T \pm z_{(1-\alpha/2)}\{\hat{p}_1(1-\hat{p}_1)/m + \hat{p}_0(1-\hat{p}_0)/(n-m)\}^{1/2} \tag{2.5}$$

and the Santner Snell (SS) exact confidence interval for a difference in binomial proportions (Santner and Snell, 1980) in a series of simulation studies. The SS confidence interval is the default exact method for a difference in binomial proportions in SAS 9.3 PROC FREQ (SAS Institute Inc., 2014). While the Wald and SS methods do not assume additivity, both assume (implicitly perhaps) that the observed data are a random sample from some larger superpopulation. In particular, the Wald and SS methods suppose the numbers of events in the treated and control groups are binomial random variables. As explained in §4 of Robins (1988), this binomial model follows from assuming either *(a)* individual potential outcomes are stochastic, Bernoulli random variables with equal mean across individuals, or *(b)* the treated and control groups constitute a random sample from some larger superpopulation. Robins argues the mean homogeneity assumption of *(a)* will usually be biologically implausible, and therefore *(b)* is implicitly being assumed whenever the binomial model is employed.

Data were simulated under three scenarios: *(i)* a randomization model, *(ii)* a randomization model under varying degrees of additivity, and *(iii)* a superpopulation model. In all simulations where $\binom{n}{m} \geq 100$, a random sample with replacement of 100 randomizations was

used to approximate permutation test p-values.

Simulations for scenario *(i)*, a randomization model, were carried out for fixed values of $n$, $m$, and $\tau$ using the following steps:

0. Potential outcomes were generated by first letting $y_j(1) = 1$ and $y_j(0) = 0$ such that $\delta_j = 1$ for individuals $j = 1, \ldots, \tau n$. Then for $j = \tau n + 1, \ldots, n$, the potential outcome $y_j(1)$ was sampled from a Bernoulli distribution with mean 0.5. Finally the potential outcomes $y_{\tau n+1}(0), \ldots, y_n(0)$ were set equal to a random permutation of $y_{\tau n+1}(1), \ldots, y_n(1)$. Generating the potential outcomes in this fashion ensured the average treatment effect equaled $\tau$.

1. Observed data were generated by randomly assigning $m$ individuals to treatment and $n-m$ individuals to control. Observed outcomes were then generated based on treatment assignment and the potential outcomes from step 0.

2. All five 95% confidence intervals (or sets) were computed for the observed data generated in step 2.

3. Steps 1-2 were repeated 1000 times.

The results for scenario *(i)* in Table 2.3 show that the permutation confidence set attained the narrowest width on average among methods that maintained nominal coverage. For all intervals (or sets) the average width decreased as $\tau$ increased for fixed $n$ and percent assigned treatment. For fixed $n$ and $\tau$, average width and coverage results were similar for 30% treatment compared to 70% treatment. The asymptotic interval was strictly narrower than the Wald interval, which is guaranteed (Robins, 1988). Coverage of the asymptotic interval tended to be substantially less than the nominal level for $\tau = 0.95$. For example, the coverage of the asymptotic interval for 70% assigned treatment and $\tau = 0.95$ was only 0.65 even when $n = 100$.

Simulations for scenario *(ii)* were carried out similar to scenario *(i)* but with varying degrees of additivity. In particular, as a measure of the amount of additivity let $\gamma = \sum_j 1\{\delta_j =$

$0\}/n$ denote the proportion of individuals where the treatment has no effect, such that $\gamma \in [0, 1]$, with the degree of additivity increasing as $\gamma \to 1$. For fixed values of $n$, $m$, and $\gamma$, simulations proceeded in the same manner as scenario *(i)* except that a different step 0 was used to generate potential outcomes. Specifically, for $j = 1, \ldots, \gamma n$, the potential outcome $y_j(1)$ was randomly sampled from a Bernoulli distribution with mean 0.5 and $y_j(0)$ was set equal to $y_j(1)$ such that $\delta_j = 0$. For individuals $j = \gamma n + 1, \ldots, (1 + \gamma)n/2$, the potential outcomes were set to $y_j(1) = 0$ and $y_j(0) = 1$ such that $\delta_j = -1$. For individuals $j = (1+\gamma)n/2+1, \ldots, n$, the potential outcomes were set to $y_j(1) = 1$ and $y_j(0) = 0$ such that $\delta_j = 1$. Generating the potential outcomes in this fashion ensured the degree of additivity equaled $\gamma$. The results for scenario *(ii)* in Table 2.4 show that the permutation confidence set again attained the narrowest width on average among methods that maintained nominal coverage. Coverage of the asymptotic interval tended to be less than the nominal level for $n \leq 60$ and $\gamma = 1$. For $n = 100$ the asymptotic interval nearly achieved the nominal level for all nine combinations of $m$ and $\gamma$.

Simulations were conducted under scenario *(iii)*, a superpopulation model, as above but with different steps 0 and 1. In particular, potential outcomes were not generated. Rather, the observed outcome data were generated by first randomly assigning $m$ of $n$ individuals to treatment. Outcomes were then independently sampled from a Bernoulli distribution with mean $p_1 = 0.5 + \Delta/2$ for individuals assigned $Z = 1$ and from a Bernoulli distribution with mean $p_0 = 0.5 - \Delta/2$ for individuals assigned $Z = 0$, where $\Delta$ was some fixed value denoting the difference in the probability of an event in the superpopulation when an individual receives treatment compared to not receiving treatment. After generating observed data, all five 95% confidence intervals (or sets) were computed. This process of data generation and interval (or set) computation was repeated 1000 times, and average interval (or set) widths and coverages were computed for the five approaches. The results for scenario *(iii)* in Table 2.5 show that the SS confidence interval was the only method to achieve nominal coverage across all simulation setups (with the exception of 30% assigned treatment at $\Delta = 0.2$ when $n = 60$). The Wald confidence interval did not reliably achieve nominal coverage with $\Delta = 0.95$, an unsurprising result given that the Wald confidence interval is known to cover poorly near

the boundary of the parameter space (Agresti and Caffo, 2000). The asymptotic confidence interval undercovered even with $n = 100$. The permutation and attributable effects confidence sets performed well, albeit with some slight undercoverage. The permutation confidence set tended to be as or more narrow than SS.

### 2.4.3 Vaccine Adherence Trial

In a study of adherence to the hepatitis B vaccine series (Seal et al., 2003), 96 injection drug users were randomized to a monetary incentive group or an outreach arm. Of the 48 individuals in the monetary incentive group, 33 were adherent, and of the 48 in the outreach arm, 11 were adherent. Using (2.3), $T = 22/48$, suggesting that 44 more individuals would have been adherent to the hepatitis B vaccine series if all 96 individuals were given monetary incentives compared to if no individuals received monetary incentives. The attributable effects confidence set is contained in the interval [0.23, 0.64]. The permutation confidence set, found using 100 re-randomizations for each hypothesis test, is contained in the interval [0.28, 0.64]. The SS, asymptotic, and Wald confidence intervals are [0.26, 0.63], [0.31, 0.60], and [0.28, 0.64] respectively. Thus for this example the permutation confidence set is the narrowest of the three exact approaches. The permutation confidence set has the same width as the Wald interval but is slightly wider than the asymptotic interval; however, unlike the Wald and asymptotic intervals, the permutation confidence set is guaranteed to cover at the nominal level.

## 2.5 Multiple Strata Designs and Observational Studies

The methods above can be extended to studies where stratified randomization is employed, i.e., individuals are randomized to treatment or control within strata. Assume that in each of $i = 1, \ldots, k$ strata, $m_i$ of $n_i$ individuals are randomized to treatment. Assume randomization is conducted independently across strata, such that there are $\prod_{i=1}^{k} \binom{n_i}{m_i}$ total possible randomizations. For stratum $i$, let $\delta_{ij}$ be the treatment effect for individual $j$ and let $\delta_i$ be the vector of treatment effects. Define $Z$ analogously for stratum $i$ such that $Z_{ij}$ is

the treatment assignment for individual $j$ and $Z_i$ is the vector of treatment assignments. The average treatment effect is $\tau = \sum_i \sum_j \delta_{ij}/n$, where $n = \sum_{i=1}^k n_i$ and where here and below $\sum_i = \sum_{i=1}^k$ and $\sum_j = \sum_{j=1}^{n_i}$.

The permutation based approach becomes computationally unwieldy in this setting. The computational burden of the permutation confidence set is based on the product of two factors. The first factor is the number of hypotheses to test. For the one stratum setting, the number of hypotheses to test is $O(n^4)$ whereas for the $k$-strata setting the number of hypotheses to test is $O(\max\{n_1, ..., n_k\}^{4k})$. The second factor is the number of permutations needed to test each hypothesis. In the one stratum problem, this number is $\binom{n}{m}$. In the $k$-strata case, the second factor is $\prod_{i=1}^k \binom{n_i}{m_i}$. Although for fixed $n = \sum_{i=1}^k n_i$ the second factor will be smaller for the $k$-strata case, the first factor in the $k$-strata case will be much larger and therefore will dominate the product. For example, suppose there are $n = 100$ individuals in $k = 4$ strata of equal sample size such that $n_1 = \cdots = n_4 = 25$; then $\max\{n_1, \ldots, n_4\}^{4k} = 25^{16} >> 100^4 = n^4$.

Given these computational challenges, the attributable effects based approach may be preferred in the multiple strata setting. To construct attributable effect based confidence sets, first note under $H_0 : \delta = \delta^0$, or equivalently that $\delta_i = \delta_i^0$ for $i = 1, \ldots, k$, the observed data can be represented in a $k$-table analogue of Table 2.2. Under this null, subtracting the attributable effect of treatment in the treated, $A_i^1(Z_i, \delta_i) = \sum_j Z_{ij}\delta_{ij}^0$, from the (1,1) cell and adding $A_i^1(Z_i, \delta_i)$ to the (1,2) cell for stratum $i = 1, \ldots, k$ will serve to fix all row and column margins in the $k$-table analogue of Table 2.2. As a result, the joint distribution of the corresponding pivotal quantities will be a product of independent hypergeometric distributions. The hypothesis $H_0 : \delta = \delta^0$ is rejected if the two-sided p-value resulting from a Cochran Mantel Haenszel exact test is sufficiently small.

As in the single stratum setting considered in Section 2, this hypothesis test can be inverted to obtain prediction sets for $A^1(Z, \delta)$ and for $A^0(Z, \delta)$. These prediction sets are considerably more difficult to find in the $k$ strata setting. As $A^1(Z, \delta) = \sum_i A_i^1(Z_i, \delta_i)$, there may be multiple combinations of $A_1^1(Z_1, \delta_1), \ldots, A_k^1(Z_k, \delta_k)$ that sum to the same value of $A^1(Z, \delta)$. Each combination producing the same $A^1(Z, \delta)$ may lead to a different p-value. A

value of $A^1(Z, \delta)$ will be included in a $1 - \alpha$ prediction set if the maximum p-value among the combinations is greater than $\alpha$. Finding the maximum p-value over the combinations of $A_1^1(Z_1, \delta_1), \ldots, A_k^1(Z_k, \delta_k)$ that sum to the same value of $A^1(Z, \delta)$ is an integer programming problem that can be solved using existing software, e.g., the R package `rgenoud` (Mebane Jr. and Sekhon, 2011). Proposition 2.1 allows for the construction of a confidence set for $\tau$ in the $k$ strata setting also.

The methods in this section may have utility in observational studies where one is willing to assume treatment selection is independent of potential outcomes conditional on some sufficient set of covariates (i.e., there are no unmeasured confounders). In this setting, an observational study can be envisaged as a stratified randomized trial performed by nature (Rosenbaum, 2002b§ 3.2), (Robins, 1988). With the strata formed by levels of the measured covariates, these methods can be employed to find exact $1 - \alpha$ confidence sets for the effect of treatment or exposure on a binary outcome.

## 2.6 Discussion

In this paper, we have presented two methods for constructing randomization based confidence sets for the average effect of a treatment on a binary outcome without assuming additivity. The first approach utilizes attributable effect sets (Rosenbaum, 2001); these sets are adjusted using a Bonferroni correction and combined to form a confidence set. The second method involves inverting a permutation test. Both methods are nonparametric, are guaranteed to yield sets that have width no greater than one, require no assumptions about random sampling from a larger population, and are exact in the sense that the probability of containing the true treatment effect is at least $1 - \alpha$. While the attributable effects method is computationally fast and the permutation method is computationally slow as $n$ increases, simulations show that permutation method has smaller average width. Based on finite population simulation results, the permutation approach is recommended over the attributable effects and asymptotic approaches for $n \leq 100$. Additional simulation results (not shown) indicate the asymptotic approach tends to provide nominal coverage for $n > 100$, although

coverage may still be less than the nominal level for extreme values of $\tau$ (e.g., $\tau \approx 1$). Extensions that allow for stratifying on categorical baseline covariates were also considered. The R package `RI2by2` is available on CRAN (Rigdon, 2014) for computing the attributable effects and permutation confidence sets as well as the asymptotic confidence interval in the one stratum setting.

There are several possible future directions to this research. For example, one future direction would be to increase the computational efficiency of the permutation based approach. Both the permutation and attributable effects based confidence sets tend to be conservative in that the empirical coverage in the simulation studies tended to be greater than the nominal level. Therefore another future research direction could explore adaptations of these two approaches which yield less conservative sets. For instance, techniques could be explored (as in Thulin (2014)) such that the average coverage equals the nominal level, although such procedures would no longer necessarily be exact.

## 2.7 Tables and Figures

Table 2.1: Cross classification of observed counts of treatment $Z$ and outcome $Y$

| | | $Y$ | | |
|---|---|---|---|---|
| | | 1 | 0 | Total |
| $Z$ | 1 | $\sum Z_j Y_j$ | $\sum Z_j(1-Y_j)$ | $m$ |
| | 0 | $\sum(1-Z_j)Y_j$ | $\sum(1-Z_j)(1-Y_j)$ | $n-m$ |
| | | $\sum Y_j$ | $\sum(1-Y_j)$ | $n$ |

Table 2.2: Cross classification of observed counts of treatment $Z$ and outcome $Y$ as a function of the potential outcomes $y_j(0)$ and the attributable effect $A^1(Z,\delta)$

| | | $Y$ | | |
|---|---|---|---|---|
| | | 1 | 0 | Total |
| $Z$ | 1 | $\sum Z_j y_j(0) + A^1(Z,\delta)$ | $\sum Z_j(1-y_j(0)) - A^1(Z,\delta)$ | $m$ |
| | 0 | $\sum(1-Z_j)y_j(0)$ | $\sum(1-Z_j)(1-y_j(0))$ | $n-m$ |
| | | $\sum y_j(0) + A^1(Z,\delta)$ | $n - \sum y_j(0) - A^1(Z,\delta)$ | $n$ |

Table 2.3: Simulation results for scenario *(i)*. Table entries give the empirical width [coverage] of 95% confidence sets or intervals, where $\tau$ is the true average treatment effect, % treatment is the percent of $n$ total individuals assigned to treatment in each experiment, Perm is the permutation confidence set, AE is the attributable effects confidence set, Asymptotic is the asymptotic confidence interval in Robins (1988), Wald is the usual large sample interval for a risk difference, and SS is the Santner-Snell Santner and Snell (1980) exact confidence interval.

| | | 30% treatment | | | 50% treatment | | | 70% treatment | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $n$ | Method | $\tau = 0.2$ | $\tau = 0.5$ | $\tau = 0.95$ | $\tau = 0.2$ | $\tau = 0.5$ | $\tau = 0.95$ | $\tau = 0.2$ | $\tau = 0.5$ | $\tau = 0.95$ |
| 20 | Perm | 0.73[1.00] | 0.70[0.99] | 0.43[1.00] | 0.75[1.00] | 0.70[1.00] | 0.34[1.00] | 0.74[1.00] | 0.70[0.99] | 0.41[1.00] |
| | AE | 0.81[1.00] | 0.76[1.00] | 0.53[1.00] | 0.88[1.00] | 0.75[1.00] | 0.52[1.00] | 0.84[1.00] | 0.76[1.00] | 0.53[1.00] |
| | Asymptotic | 0.80[0.96] | 0.68[1.00] | 0.12[0.30] | 0.76[0.96] | 0.60[0.98] | 0.11[0.49] | 0.82[0.97] | 0.67[0.91] | 0.10[0.70] |
| | Wald | 0.87[0.97] | 0.79[0.96] | 0.14[0.30] | 0.83[0.99] | 0.73[0.98] | 0.14[0.49] | 0.89[0.98] | 0.79[0.94] | 0.14[0.70] |
| | SS | 0.91[1.00] | 0.83[1.00] | 0.52[1.00] | 0.89[1.00] | 0.80[1.00] | 0.41[1.00] | 0.91[1.00] | 0.83[1.00] | 0.49[1.00] |
| 40 | Perm | 0.59[1.00] | 0.53[1.00] | 0.24[1.00] | 0.58[1.00] | 0.51[1.00] | 0.20[1.00] | 0.60[1.00] | 0.52[1.00] | 0.26[1.00] |
| | AE | 0.68[1.00] | 0.61[1.00] | 0.41[1.00] | 0.67[1.00] | 0.58[1.00] | 0.30[1.00] | 0.68[1.00] | 0.61[1.00] | 0.32[1.00] |
| | Asymptotic | 0.60[0.97] | 0.49[0.97] | 0.08[0.91] | 0.55[0.98] | 0.42[1.00] | 0.11[0.76] | 0.60[0.96] | 0.47[0.97] | 0.11[0.79] |
| | Wald | 0.65[0.98] | 0.58[0.99] | 0.12[0.91] | 0.60[1.00] | 0.52[1.00] | 0.13[0.76] | 0.65[0.97] | 0.56[0.98] | 0.14[0.79] |
| | SS | 0.66[1.00] | 0.60[1.00] | 0.29[1.00] | 0.63[1.00] | 0.57[1.00] | 0.26[1.00] | 0.66[1.00] | 0.61[1.00] | 0.31[1.00] |
| 60 | Perm | 0.51[0.99] | 0.43[1.00] | 0.19[1.00] | 0.50[1.00] | 0.42[1.00] | 0.15[1.00] | 0.52[1.00] | 0.44[1.00] | 0.19[1.00] |
| | AE | 0.57[1.00] | 0.49[1.00] | 0.24[1.00] | 0.57[1.00] | 0.50[1.00] | 0.23[1.00] | 0.57[1.00] | 0.52[1.00] | 0.24[1.00] |
| | Asymptotic | 0.50[0.98] | 0.37[0.98] | 0.10[0.78] | 0.45[0.98] | 0.35[0.98] | 0.10[1.00] | 0.50[0.99] | 0.41[0.98] | 0.09[0.64] |
| | Wald | 0.53[0.98] | 0.45[1.00] | 0.13[1.00] | 0.49[1.00] | 0.43[0.99] | 0.13[1.00] | 0.53[0.99] | 0.48[0.99] | 0.12[0.95] |
| | SS | 0.54[0.99] | 0.50[1.00] | 0.23[1.00] | 0.52[1.00] | 0.46[1.00] | 0.20[1.00] | 0.54[1.00] | 0.49[1.00] | 0.23[1.00] |
| 100 | Perm | 0.42[1.00] | 0.35[1.00] | 0.14[1.00] | 0.40[1.00] | 0.33[1.00] | 0.11[1.00] | 0.42[1.00] | 0.36[1.00] | 0.14[0.99] |
| | AE | 0.47[1.00] | 0.41[1.00] | 0.18[1.00] | 0.45[1.00] | 0.39[1.00] | 0.17[1.00] | 0.46[1.00] | 0.42[1.00] | 0.18[1.00] |
| | Asymptotic | 0.38[0.98] | 0.31[0.98] | 0.08[0.72] | 0.35[0.99] | 0.27[0.98] | 0.09[0.88] | 0.39[0.99] | 0.31[0.98] | 0.09[0.65] |
| | Wald | 0.41[0.99] | 0.37[0.99] | 0.11[0.96] | 0.38[1.00] | 0.33[1.00] | 0.11[1.00] | 0.42[0.99] | 0.37[1.00] | 0.11[0.82] |
| | SS | 0.42[0.99] | 0.38[1.00] | 0.17[1.00] | 0.40[1.00] | 0.35[1.00] | 0.15[1.00] | 0.42[0.99] | 0.38[1.00] | 0.17[1.00] |

Table 2.4: Simulation results for scenario *(ii)*. Table entries give the empirical width [coverage] of 95% confidence sets or intervals, where $\gamma$ is the degree of additivity, % treatment is the percent of $n$ total individuals assigned to treatment in each experiment, Perm is the permutation confidence set, AE is the attributable effects confidence set, Asymptotic is the asymptotic confidence interval in Robins (1988), Wald is the usual large sample interval for a risk difference, and SS is the Santner-Snell Santner and Snell (1980) exact confidence interval.

| | | 30% treatment | | | 50% treatment | | | 70% treatment | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $n$ | Method | $\gamma = 0.2$ | $\gamma = 0.8$ | $\gamma = 1$ | $\gamma = 0.2$ | $\gamma = 0.8$ | $\gamma = 1$ | $\gamma = 0.2$ | $\gamma = 0.8$ | $\gamma = 1$ |
| 20 | Perm | 0.73[1.00] | 0.74[0.99] | 0.72[0.99] | 0.76[1.00] | 0.76[0.99] | 0.66[0.98] | 0.73[1.00] | 0.73[0.99] | 0.67[0.98] |
| | AE | 0.88[1.00] | 0.86[1.00] | 0.84[1.00] | 0.91[1.00] | 0.91[1.00] | 0.77[0.99] | 0.88[1.00] | 0.83[1.00] | 0.80[0.99] |
| | Asymptotic | 0.86[0.99] | 0.83[0.95] | 0.81[0.89] | 0.82[1.00] | 0.80[0.92] | 0.69[0.93] | 0.86[0.99] | 0.80[0.93] | 0.73[0.92] |
| | Wald | 0.91[0.99] | 0.89[0.95] | 0.86[0.95] | 0.85[1.00] | 0.85[0.96] | 0.75[0.94] | 0.91[0.99] | 0.86[0.93] | 0.78[0.93] |
| | SS | 0.94[1.00] | 0.93[0.97] | 0.93[0.96] | 0.91[1.00] | 0.90[0.98] | 0.78[0.99] | 0.94[1.00] | 0.93[0.99] | 0.78[0.98] |
| 40 | Perm | 0.60[1.00] | 0.60[0.98] | 0.59[0.98] | 0.59[1.00] | 0.59[0.99] | 0.54[0.97] | 0.60[1.00] | 0.60[0.99] | 0.56[0.98] |
| | AE | 0.69[1.00] | 0.69[0.99] | 0.67[0.99] | 0.68[1.00] | 0.67[1.00] | 0.64[0.99] | 0.68[1.00] | 0.69[0.99] | 0.64[1.00] |
| | Asymptotic | 0.64[0.99] | 0.62[0.92] | 0.61[0.90] | 0.59[1.00] | 0.57[0.94] | 0.55[0.92] | 0.63[0.99] | 0.62[0.91] | 0.57[0.92] |
| | Wald | 0.66[1.00] | 0.66[0.95] | 0.65[0.93] | 0.61[1.00] | 0.60[0.94] | 0.58[0.95] | 0.65[0.99] | 0.66[0.95] | 0.60[0.94] |
| | SS | 0.67[1.00] | 0.67[0.97] | 0.65[0.96] | 0.65[1.00] | 0.64[0.98] | 0.58[0.97] | 0.67[1.00] | 0.67[0.98] | 0.60[0.97] |
| 60 | Perm | 0.52[1.00] | 0.52[0.99] | 0.52[0.98] | 0.51[1.00] | 0.50[0.98] | 0.47[0.98] | 0.52[1.00] | 0.52[0.99] | 0.49[0.97] |
| | AE | 0.59[1.00] | 0.58[1.00] | 0.58[0.99] | 0.58[1.00] | 0.56[1.00] | 0.53[0.99] | 0.59[1.00] | 0.59[0.99] | 0.56[0.98] |
| | Asymptotic | 0.53[1.00] | 0.52[0.94] | 0.51[0.90] | 0.49[1.00] | 0.47[0.94] | 0.47[0.90] | 0.53[1.00] | 0.52[0.93] | 0.49[0.90] |
| | Wald | 0.54[1.00] | 0.54[0.96] | 0.53[0.94] | 0.50[1.00] | 0.49[0.96] | 0.49[0.95] | 0.54[1.00] | 0.54[0.95] | 0.51[0.93] |
| | SS | 0.55[1.00] | 0.55[0.97] | 0.54[0.97] | 0.53[1.00] | 0.52[0.99] | 0.49[0.96] | 0.55[1.00] | 0.55[0.96] | 0.51[0.96] |
| 100 | Perm | 0.43[1.00] | 0.43[0.99] | 0.42[0.97] | 0.41[1.00] | 0.41[0.99] | 0.39[0.97] | 0.43[1.00] | 0.43[0.98] | 0.41[0.97] |
| | AE | 0.47[1.00] | 0.47[1.00] | 0.47[0.98] | 0.45[1.00] | 0.45[0.99] | 0.43[0.98] | 0.47[1.00] | 0.47[0.99] | 0.44[0.98] |
| | Asymptotic | 0.41[1.00] | 0.41[0.96] | 0.41[0.92] | 0.38[1.00] | 0.38[0.96] | 0.37[0.94] | 0.42[1.00] | 0.41[0.94] | 0.39[0.94] |
| | Wald | 0.42[1.00] | 0.42[0.97] | 0.42[0.94] | 0.39[1.00] | 0.39[0.97] | 0.38[0.95] | 0.42[1.00] | 0.42[0.96] | 0.40[0.94] |
| | SS | 0.43[1.00] | 0.43[0.98] | 0.42[0.95] | 0.41[1.00] | 0.41[0.98] | 0.39[0.96] | 0.43[1.00] | 0.43[0.97] | 0.41[0.96] |

Table 2.5: Simulation results for scenario *(iii)*. Table entries give the empirical width [coverage] of 95% confidence sets or intervals, where $\Delta$ is the true difference in binomial proportions, % treatment is the percent of $n$ total individuals assigned to treatment in each experiment, Perm is the permutation confidence set, AE is the attributable effects confidence set, Asymptotic is the asymptotic confidence interval in Robins (1988), Wald is the usual large sample interval for a difference in binomial proportions, and SS is the Santner-Snell Santner and Snell (1980) exact confidence interval.

| | | 30% treatment | | | 50% treatment | | | 70% treatment | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $n$ | Method | $\Delta = 0.2$ | $\Delta = 0.5$ | $\Delta = 0.95$ | $\Delta = 0.2$ | $\Delta = 0.5$ | $\Delta = 0.95$ | $\Delta = 0.2$ | $\Delta = 0.5$ | $\Delta = 0.95$ |
| 20 | Perm | 0.73[0.95] | 0.68[0.95] | 0.42[0.90] | 0.74[0.96] | 0.69[0.94] | 0.34[0.92] | 0.73[0.95] | 0.67[0.94] | 0.42[0.91] |
| | AE | 0.83[0.98] | 0.75[0.98] | 0.53[0.90] | 0.85[0.98] | 0.75[0.98] | 0.53[0.92] | 0.83[0.99] | 0.75[0.98] | 0.53[0.91] |
| | Asymptotic | 0.78[0.84] | 0.63[0.82] | 0.11[0.39] | 0.73[0.86] | 0.59[0.89] | 0.11[0.41] | 0.78[0.85] | 0.61[0.78] | 0.11[0.38] |
| | Wald | 0.86[0.90] | 0.75[0.89] | 0.13[0.40] | 0.81[0.92] | 0.71[0.90] | 0.14[0.42] | 0.86[0.91] | 0.73[0.85] | 0.13[0.39] |
| | SS | 0.90[0.96] | 0.82[0.98] | 0.50[1.00] | 0.87[0.97] | 0.79[0.97] | 0.41[0.99] | 0.91[0.96] | 0.81[0.98] | 0.50[1.00] |
| 40 | Perm | 0.59[0.95] | 0.52[0.94] | 0.26[0.91] | 0.58[0.97] | 0.51[0.96] | 0.20[0.92] | 0.59[0.95] | 0.52[0.95] | 0.26[0.94] |
| | AE | 0.67[0.98] | 0.60[0.98] | 0.32[0.91] | 0.66[0.98] | 0.58[0.98] | 0.30[0.92] | 0.67[0.98] | 0.60[0.98] | 0.32[0.94] |
| | Asymptotic | 0.59[0.90] | 0.47[0.86] | 0.10[0.61] | 0.54[0.88] | 0.43[0.88] | 0.10[0.61] | 0.58[0.88] | 0.47[0.87] | 0.10[0.66] |
| | Wald | 0.64[0.92] | 0.56[0.90] | 0.12[0.62] | 0.59[0.95] | 0.52[0.94] | 0.12[0.62] | 0.63[0.90] | 0.56[0.91] | 0.13[0.66] |
| | SS | 0.66[0.99] | 0.60[0.96] | 0.31[0.99] | 0.63[0.97] | 0.56[0.98] | 0.25[0.98] | 0.65[0.95] | 0.59[0.97] | 0.31[0.99] |
| 60 | Perm | 0.51[0.94] | 0.44[0.94] | 0.20[0.92] | 0.49[0.96] | 0.42[0.96] | 0.15[0.93] | 0.51[0.96] | 0.44[0.95] | 0.20[0.93] |
| | AE | 0.57[0.96] | 0.51[0.97] | 0.24[0.92] | 0.56[0.98] | 0.50[0.99] | 0.23[0.93] | 0.57[0.97] | 0.50[0.98] | 0.24[0.93] |
| | Asymptotic | 0.49[0.91] | 0.40[0.87] | 0.10[0.55] | 0.44[0.89] | 0.35[0.90] | 0.09[0.74] | 0.49[0.91] | 0.39[0.89] | 0.09[0.54] |
| | Wald | 0.53[0.93] | 0.47[0.92] | 0.12[0.80] | 0.49[0.93] | 0.43[0.95] | 0.11[0.76] | 0.53[0.92] | 0.46[0.93] | 0.12[0.78] |
| | SS | 0.54[0.94] | 0.49[0.96] | 0.24[0.99] | 0.51[0.97] | 0.46[0.97] | 0.19[0.81] | 0.54[0.95] | 0.49[0.96] | 0.23[0.99] |
| 100 | Perm | 0.42[0.95] | 0.35[0.94] | 0.14[0.93] | 0.39[0.96] | 0.33[0.93] | 0.11[0.95] | 0.42[0.96] | 0.35[0.93] | 0.14[0.93] |
| | AE | 0.46[0.97] | 0.41[0.98] | 0.18[0.96] | 0.45[0.98] | 0.39[0.98] | 0.17[0.96] | 0.46[0.98] | 0.41[0.97] | 0.18[0.96] |
| | Asymptotic | 0.38[0.90] | 0.31[0.88] | 0.08[0.64] | 0.35[0.90] | 0.27[0.88] | 0.08[0.68] | 0.38[0.92] | 0.31[0.87] | 0.08[0.64] |
| | Wald | 0.41[0.92] | 0.36[0.93] | 0.11[0.79] | 0.38[0.94] | 0.34[0.94] | 0.10[0.91] | 0.41[0.94] | 0.36[0.93] | 0.11[0.79] |
| | SS | 0.42[0.95] | 0.38[0.96] | 0.17[0.98] | 0.40[0.95] | 0.35[0.95] | 0.14[0.98] | 0.42[0.96] | 0.38[0.96] | 0.17[0.99] |

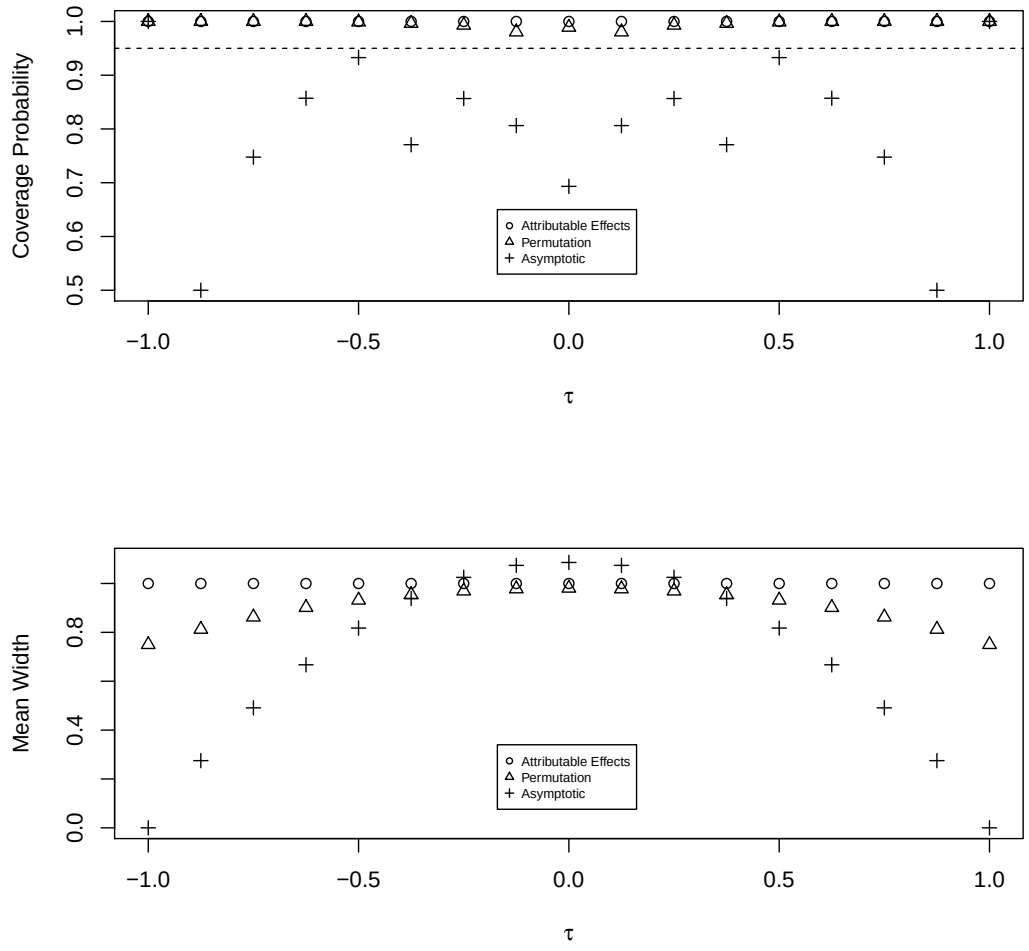


Figure 2.1: Coverage probability (top) and average width (bottom) of the attributable effects and permutation test based confidence sets and the asymptotic confidence interval for the average treatment effect $\tau$.

# CHAPTER 3: EXACT CONFIDENCE INTERVALS IN THE PRESENCE OF INTERFERENCE

## 3.1 Introduction

In a randomized experiment, it is commonly assumed that an individual only has two potential outcomes: an outcome on control, and an outcome on treatment. That an individual has only two potential outcomes assumes no interference (Cox, 1958) between individuals, i.e., an individual's potential outcomes are unaffected by the treatment assignment of any other individual in the study. There are many settings where this assumption of no interference is clearly violated (Hong and Raudenbush, 2006; Sobel, 2006; Rosenbaum, 2007).

Partial interference holds when individuals can be partitioned into groups such that there is no interference between individuals in different groups. In settings where partial interference holds, two-stage randomized experiments have been suggested as a study design for drawing inference about treatment (i.e., causal) effects. Two-stage randomized experiments proceed by *(i)* randomizing groups to treatment strategies and *(ii)* randomizing individuals within groups to different treatments based on the treatment strategy assigned to their group in stage *(i)*. Two-stage randomized experiments are found in many fields of study, e.g., infectious diseases (Baird et al., 2012), medicine (Borm et al., 2005), economics (Duflo and Saez, 2003), and political science (Ichino and Schündeln, 2012; Sinclair et al., 2012). Building upon ideas in Halloran et al. (1991), Hudgens and Halloran (2008) defined and derived unbiased estimators for the direct, indirect, total, and overall effects of treatment in a two-stage randomized experiment assuming partial interference. Liu and Hudgens (2014) showed that Wald-type confidence intervals based on these estimators perform well when the number of groups is large; however, often the number of groups may not be large enough. For example, Moulton et al. (2001) describe a group-randomized vaccine trial involving approximately 9,000 individuals

but only 38 groups. Tchetgen Tchetgen and VanderWeele (2012), henceforth TV, proposed exact confidence intervals using the Hoeffding inequality for these four effects in a two-stage randomized experiment with partial interference. Unfortunately, as will be shown below, the TV intervals can be very wide and conservative.

In this paper, we propose different exact confidence intervals based on inverting exact hypothesis tests that tend to be less conservative than TV. The remainder of the paper is organized as follows. In §2, treatment effects in the presence of interference are defined and existing inferential results are reviewed. In §3, the assumption of stratified interference is presented and bounds are derived for the causal effects under this assumption. In §4 the proposed new exact confidence intervals are described by inverting certain permutation tests. §5 concludes with a simulation study comparing the TV, asymptotic, and new exact confidence intervals.

## 3.2 Preliminaries

### 3.2.1 Estimands

Consider a finite population of $N$ individuals partitioned into $k$ groups with $n_i$ individuals in group $i$ for $i = 1, \ldots, k$. Assume partial interference, i.e., there is no interference between individuals in different groups. Consider a two-stage randomized experiment wherein $h$ of $k$ groups are assigned to strategy $\alpha_1$ and $k - h$ are assigned to $\alpha_0$ in the first stage, where strategy $\alpha_s$ specifies that $m_i^s$ of $n_i$ individuals will receive treatment. For example, strategy $\alpha_0$ might entail assigning (approximately) 1/3 of individuals within a group to treatment whereas strategy $\alpha_1$ might entail assigning (approximately) 2/3 of individuals within a group to treatment (see TV for further discussion about different types of treatment allocation strategies). Let $S_i = 1$ if group $i$ is randomized to $\alpha_1$ and 0 otherwise so that $\Pr[S_i = 1] = h/k$. In the second stage, individuals will be randomized to treatment conditional on group assignment in the first stage. Let $Z_{ij} = 1$ if individual $j$ in group $i$ is assigned treatment and 0 otherwise. Let $Z_i = (Z_{i1}, \ldots, Z_{in_i})$ be the random vector of treatment assignments

for group $i$ taking on values $z_i \in \mathcal{R}(n_i, m_i^s)$, the set of all vectors of length $n_i$ composed of $m_i^s$ elements equal to 1 and $n_i - m_i^s$ elements equal to 0. Additionally, let $Z_{i(j)}$ denote the random vector of treatment assignments in group $i$ excluding individual $j$ taking on values $z_{i(j)} \in \mathcal{R}(n_i - 1, m_i^s - z_{ij})$.

Let $y_{ij}(z_i)$ be the binary potential outcome for individual $j$ in group $i$ when group $i$ receives treatment vector $z_i$. A randomization inference framework is adopted wherein potential outcomes are fixed features of the finite population of $N$ individuals and only treatment assignments $S$ and $Z$ are random (as in Sobel (2006); Rosenbaum (2007); Hudgens and Halloran (2008)). Define the average potential outcome for individual $j$ in group $i$ on treatment $z = 0, 1$ under strategy $\alpha_s$ as

$$\bar{y}_{ij}(z; \alpha_s) \equiv \sum_{\omega \in \mathcal{R}(n_i - 1, m_i^s - z)} y_{ij}(z_{ij} = z, z_{i(j)} = \omega) \Pr(Z_{i(j)} = \omega | Z_{ij} = z; S_i = s) \tag{3.1}$$

where $\Pr(Z_{i(j)} = \omega | Z_{ij} = z; S_i = s) = \binom{n_i - 1}{m_i^s - z}^{-1}$. Henceforth, let $\sum_i = \sum_{i=1}^k$ and $\sum_j = \sum_{j=1}^{n_i}$. For treatment $z$ under strategy $\alpha_s$ define the group average potential outcome as $\bar{y}_i(z; \alpha_s) \equiv n_i^{-1} \sum_j \bar{y}_{ij}(z; \alpha_s)$, and the population average potential outcome as $\bar{y}(z; \alpha_s) \equiv k^{-1} \sum_i \bar{y}_i(z; \alpha_s)$. Define the average potential outcome for individual $j$ in group $i$ under strategy $\alpha_s$ as

$$\bar{y}_{ij}(\alpha_s) \equiv \sum_{\omega \in \mathcal{R}(n_i, m_i^s)} y_{ij}(z_i = \omega) \Pr(Z_i = \omega; S_i = s), \tag{3.2}$$

the group average potential outcome as $\bar{y}_i(\alpha_s) \equiv n_i^{-1} \sum_j \bar{y}_{ij}(\alpha_s)$, and the population average potential outcome as $\bar{y}(\alpha_s) \equiv k_i^{-1} \sum_i \bar{y}_i(\alpha_s)$. Define the direct effect of treatment for strategy $\alpha_s$ as $DE(\alpha_s) = \bar{y}(0; \alpha_s) - \bar{y}(1; \alpha_s)$, the indirect effect of $\alpha_0$ versus $\alpha_1$ as $IE(\alpha_0, \alpha_1) = \bar{y}(0; \alpha_0) - \bar{y}(0; \alpha_1)$, the total effect as $TE(\alpha_0, \alpha_1) = \bar{y}(0; \alpha_0) - \bar{y}(1; \alpha_1)$, and the overall effect of $\alpha_0$ versus $\alpha_1$ as $OE(\alpha_0, \alpha_1) = \bar{y}(\alpha_0) - \bar{y}(\alpha_1)$; see Hudgens and Halloran (2008) and TV for additional discussion regarding these effects.

### 3.2.2 Existing Inferential Results

Hudgens and Halloran (2008) derived unbiased estimators for all population average potential outcomes, and thus for the four causal effects. Noting that $\Pr[S_i = s]$ and $\Pr[Z_{ij} = z|S_i = s]$ are known by design, the estimator

$$\hat{y}(z;\alpha_s) = k^{-1}\sum_i \frac{1\{S_i = s\}\hat{y}_i(z;\alpha_s)}{\Pr[S_i = s]} \tag{3.3}$$

where $\hat{y}_i(z;\alpha_s) = n_i^{-1}\sum_j 1\{Z_{ij} = z\}y_{ij}(Z_{ij})/\Pr[Z_{ij} = z|S_i = s]$ is unbiased for $\bar{y}(z;\alpha_s)$. Additionally, the estimator

$$\hat{y}(\alpha_s) = k^{-1}\sum_i \frac{1(S_i = s)n_i^{-1}\sum_j y_{ij}(Z_{ij})}{\Pr[S_i = s]} \tag{3.4}$$

is unbiased for $\bar{y}(\alpha_s)$. Unbiased estimators for the effects of interest follow immediately: $\widehat{DE}(\alpha_s) = \hat{y}(0;\alpha_s) - \hat{y}(1;\alpha_s)$, $\widehat{IE}(\alpha_0,\alpha_1) = \hat{y}(0;\alpha_0) - \hat{y}(0;\alpha_1)$, $\widehat{TE}(\alpha_0,\alpha_1) = \hat{y}(0;\alpha_0) - \hat{y}(1;\alpha_1)$, and $\widehat{OE}(\alpha_0,\alpha_1) = \hat{y}(\alpha_0) - \hat{y}(\alpha_1)$.

TV proposed exact confidence intervals based on the Hoeffding inequality for the effects of interest in a two-stage randomized experiment where partial interference is assumed. In particular, for any $\gamma \in \{0,1\}$, $\widehat{DE}(\alpha_s) \pm \epsilon_D^*(\gamma,\alpha_s,q_s,k)$ is a $1-\gamma$ exact confidence interval for $DE(\alpha_s)$ where $\epsilon_D^*(\gamma,\alpha_s,q_s,k)$ is given in equation (17) of TV for $s = 0,1$. Additionally, $\widehat{IE}(\alpha_0,\alpha_1) \pm \epsilon^*(\gamma,\alpha_0,q_0,\alpha_1,q_1,k)$, $\widehat{TE}(\alpha_0,\alpha_1) \pm \epsilon^*(\gamma,\alpha_0,q_0,\alpha_1,q_1,k)$, and $\widehat{OE}(\alpha_0,\alpha_1) \pm \epsilon^*(\gamma,\alpha_0,q_0,\alpha_1,q_1,k)$ are all $1-\gamma$ exact confidence intervals for their target parameters where $\epsilon^*(\gamma,\alpha_0,q_0,\alpha_1,q_1,k)$ is given in Theorem 3 of TV.

Liu and Hudgens (2014) examined conditions under which Wald-type intervals $\widehat{DE}(\alpha_s) \pm z_{(1-\gamma/2)}\{\widehat{var}(\widehat{DE}(\alpha_s))\}^{1/2}$ and Chebyshev-type intervals $\widehat{DE}(\alpha_s) \pm \{\widehat{var}(\widehat{DE}(\alpha_s))/\gamma\}^{1/2}$ are valid, large sample confidence intervals for $DE(\alpha_s)$, where $z_{(1-\gamma/2)}$ is the $1-\gamma/2$ quantile for the standard normal distribution and $\widehat{var}(\widehat{DE}(\alpha_s))$ is an estimator of the variance of $\widehat{DE}(\alpha_s)$ for $s = 0,1$. They also considered Wald and Chebyshev-type confidence intervals for the indirect, total, and overall effects.

### 3.3 Bounds Under Stratified Interference

Exact randomization based inference about the four effects is challenging without further assumptions as the experiment reveals only $N$ of the $\sum_i \sum_j \left\{ \binom{n_i}{m_i^0} + \binom{n_i}{m_i^1} \right\}$ total potential outcomes. One such additional assumption is stratified interference (Hudgens and Halloran, 2008), which assumes that individual $j$ in group $i$ has the same potential outcome when assigned control or treatment as long as a fixed number of other individuals in group $i$ are assigned treatment, i.e.,

$$y_{ij}(z_i) = y_{ij}(z_i') \text{ for all } z_i, z_i' \in \mathcal{R}(n_i, m_i^s) \text{ such that } z_{ij} = z_{ij}'. \tag{3.5}$$

Under (3.5), individual $j$ in group $i$ only has four potential outcomes, which we denote by $y_{ij}(z; \alpha_s)$ for $z, s = 0, 1$, so that the experiment reveals the observed outcome $Y_{ij} = \sum_{z,s=0,1} 1\{Z_{ij} = z; S_i = s\} y_{ij}(z; \alpha_s)$ for each individual and thus $N$ of the $4N$ total potential outcomes. Furthermore, (3.5) implies that $\bar{y}_{ij}(z; \alpha_s) = y_{ij}(z; \alpha_s)$, and that $\bar{y}_{ij}(\alpha_s) = w_i^s y_{ij}(1; \alpha_s) + (1 - w_i^s) y_{ij}(0; \alpha_s) \equiv y_{ij}(\alpha_s)$ where $w_i^s = \Pr[Z_{ij} = 1 | S_i = s] = m_i^s / n_i$.

Under (3.5), the observed data form bounded sets for all effects contained in the interval $[-1, 1]$. The bounded sets have widths less than two where here and in the sequel the width of a set is defined to be the difference between its maximum and minimum values. Consider $DE(\alpha_0) = k^{-1} \sum_i n_i^{-1} \sum_j \{y_{ij}(0; \alpha_0) - y_{ij}(1; \alpha_0)\}$ for illustration. For the $\sum_i \sum_j (1 - S_i)(1 - Z_{ij})$ individuals with $S_i = Z_{ij} = 0$, $y_{ij}(0; \alpha_0)$ is revealed; however, for the $N - \sum_i \sum_j (1 - S_i)(1 - Z_{ij})$ individuals with $S_i = 1$ or $Z_{ij} = 1$, $y_{ij}(0; \alpha_0)$ is missing and only known to be 0 or 1. Let $\vec{y}(z; \alpha_s)$ be the $N$-dimensional vector of potential outcomes for treatment $z$ under strategy $\alpha_s$. Under (3.5), a lower bound for $DE(\alpha_0)$ is found by filling in all missing potential outcomes in $\vec{y}(0; \alpha_0)$ as 0 and all missing potential outcomes in $\vec{y}(1; \alpha_0)$ as 1. An upper bound for $DE(\alpha_0)$ is found by filling in all missing potential outcomes in $\vec{y}(0; \alpha_0)$ as 1 and all missing potential outcomes in $\vec{y}(1; \alpha_0)$ as 0. Simple algebra shows that width of the bounded set for $DE(\alpha_0)$ is equal to $2 - (k - h)/k$. The width of this bounded set approaches 1 as $(k - h)/k \to 1$, i.e., as more groups are randomized to $\alpha_0$.

Similar logic leads to bounds for the other effects. The width of the bounded set for $DE(\alpha_1)$ is equal to $2 - h/k$ which approaches 1 as $h/k \to 1$. The width of the bounded set for $IE(\alpha_0, \alpha_1)$ is equal to $2 - k^{-1} \sum_i n_i^{-1} \sum_j (1 - Z_{ij})$ which approaches 1 as the proportion of individuals assigned $Z_{ij} = 0$ approaches 1. The width of the bounded set for $TE(\alpha_0, \alpha_1)$ is equal to $2 - k^{-1} \sum_i n_i^{-1} \{(1 - S_i) \sum_j (1 - Z_{ij}) + S_i \sum_j Z_{ij}\}$ which approaches 1 as the proportion of individuals with $S_i = Z_{ij} = 0$ or $S_i = Z_{ij} = 1$ approaches 1. Lower and upper bounds for $OE(\alpha_0, \alpha_1)$ can be derived similarly but the corresponding width does not have a simple closed form.

## 3.4 EIT Confidence Intervals

In addition to leading to unbiased estimators and bounds, the observed data can be used to form $1 - \gamma$ confidence sets for the four effects. The confidence sets are formed by inverting hypothesis tests about the potential outcomes that define the effect of interest. This section is divided into two parts: §3.4.1 outlines how the confidence sets are formed and §3.4.2 presents a computationally feasible algorithm for constructing an interval that contains the exact confidence set. Henceforth this interval is referred to as the exact inverted test (EIT).

### 3.4.1 An Exact Confidence Set

The methods to follow can be generalized to any effect, so consider $DE(\alpha_0)$. Inference about $DE(\alpha_0)$ concerns the vectors $\vec{y}(0; \alpha_0)$ and $\vec{y}(1; \alpha_0)$, which are partially revealed by the experiment. A hypothesis about these vectors is considered sharp if it completely fills in the potential outcomes not revealed by the experiment. A sharp null $H_0 : \vec{y}(0; \alpha_0) = \vec{y}^0(0; \alpha_0), \vec{y}(1; \alpha_0) = \vec{y}^0(1; \alpha_0)$ maps to a value of $DE(\alpha_0)$, which we denote $DE^0(\alpha_0)$. Only sharp null hypotheses that are compatible with the observed data need to be tested as other sharp nulls can be rejected with zero probability of making a type I error. Thus for each sharp null to be tested, the implied null value $DE^0(\alpha_0)$ will be a member of the bounded set derived in §3. There are $B_1 = 2^{\sum_i (1 - S_i) n_i} 4^{\sum_i S_i n_i}$ sharp null hypotheses to test, as individuals with $S_i = 0$ have only one missing potential outcome with two possible values

$\{0,1\}$, and individuals with $S_i = 1$ have two missing potential outcomes with four possible values $\{0,1\} \times \{0,1\}$.

After filling in the missing potential outcomes under $H_0$, the null distribution of the test statistic $\widehat{DE}(\alpha_0)$ can be found by computing the statistic, denoted by $\widehat{DE}_c(\alpha_0)$, for each of the $c = 1, \ldots, C_1$ possible experiments under $H_0$, where $C_1 = \sum_{S \in \mathcal{S}} \prod_{i=1}^{k} \binom{n_i}{m_i^1}^{S_i} \binom{n_i}{m_i^0}^{(1-S_i)}$ and $\mathcal{S}$ is the set of all possible values of the vector $S$ such that $|\mathcal{S}| = \binom{k}{h}$. A two-sided p-value to test $H_0$ is given by $p_0 = \sum_{c=1}^{C_1} 1\{|\widehat{DE}_c(\alpha_0) - DE^0(\alpha_0)| \geq |\widehat{DE}(\alpha_0) - DE^0(\alpha_0)|\}/C_1$. If $p_0 < \gamma$, $H_0$ is rejected. Note $p_0$ is a function of the null hypothesis vectors $\vec{y}^0(0;\alpha_0)$ and $\vec{y}^0(1;\alpha_0)$. Let $p(DE^0(\alpha_0))$ denote the set of all $p_0$ which are functions of compatible vectors $\vec{y}^0(0;\alpha_0)$ and $\vec{y}^0(1;\alpha_0)$ that map to $DE^0(\alpha_0)$. A $1-\gamma$ confidence set for $DE(\alpha_0)$ is $\{DE^0(\alpha_0) : \max\{p(DE^0(\alpha_0))\} \geq \gamma\}$. P-values, and thus confidence sets, can be found in an analogous manner for the other effects.

### 3.4.2 A Computationally Feasible Algorithm

Finding the exact confidence set for $DE(\alpha_0)$ described above entails testing $B_1$ hypotheses, where each hypothesis test involves $C_1$ randomizations. As $N$ becomes large, the computational time necessary to perform $B_1 \times C_1$ operations grows exponentially. For illustration of the problem, consider two examples in which $h = 1$ of $k = 1$ groups are assigned $\alpha_0$, in which $m_1^0 = 10$ of $n_1 = 20$ individuals are randomized to treatment such that $B_1 = 2^{20}$ and $C_1 = \binom{20}{10} = 184,756$. Suppose there are two cases of observed data: (a) 5 of 10 unexposed experienced an event, and 5 of 10 exposed experienced an event, and (b) 8 of 10 unexposed experienced an event and 2 of 10 exposed experienced an event. Figure 4.1 displays a plot of $DE^0(\alpha_0)$ versus $p(DE^0(\alpha_0))$ for both examples. The bounded set and 95% exact confidence set for $DE^0(\alpha_0)$, respectively are $\{-0.5, -0.45, \ldots, 0.45, 0.5\}$ and $\{-0.35, -0.3, \ldots, 0.3, 0.35\}$ in (a) and $\{-0.2, -0.15, \ldots, 0.75, 0.8\}$ and $\{0.15, 0.2, \ldots, 0.75, 0.8\}$ in (b).

To save computational time in finding the confidence sets, $B_2 < B_1$ of the sharp null hypotheses can be tested, and a random sample of $C_2 < C_1$ randomizations can be used to compute the p-value for each sharp null. The lack of symmetry in Figure 1 (b) suggests

targeting the lower and upper limit of the confidence set separately. Consider the following targeting algorithm for the lower bound of a confidence set for $DE(\alpha_0)$ where $\widehat{DE}(\alpha_0)_l$ denotes the lower bound for $DE(\alpha_0)$, and $\hat{y}(z;\alpha_0)_l$ and $\hat{y}(z;\alpha_0)_u$ denote the lower and upper bounds, respectively, for $\bar{y}(z;\alpha_0)$. An analogous algorithm can be used to target the upper limit of the confidence set for $DE(\alpha_0)$.

1. Test the unique sharp null about $\vec{y}(0;\alpha_0)$ and $\vec{y}(1;\alpha_0)$ that maps to $\widehat{DE}(\alpha_0)_l$. If the corresponding p-value $p_0 \geq \gamma$, let $\widehat{DE}(\alpha_0)_l$ be the lower limit of the confidence set and do not proceed. Otherwise, let $l = \widehat{DE}(\alpha_0)$ and let $p_l = 1 - 1/B_2$. Let $\mathcal{L} = \{\widehat{DE}(\alpha_0)_l\}$ and $\mathcal{P} = \{p_0\}$.

2. Fill in the missingness in $\vec{y}(0;\alpha_0)$ with samples from a Bernoulli distribution with mean $f(\{\hat{y}(0;\alpha_0)_l + \hat{y}(1;\alpha_0)_u + q_{p_l}(\widehat{DE}(\alpha_0)_l, l)\}/2)$ and fill in the missingness in $\vec{y}(1;\alpha_0)$ with samples from a Bernoulli distribution with mean $f(\{\hat{y}(0;\alpha_0)_l + \hat{y}(1;\alpha_0)_u - q_{p_l}(\widehat{DE}(\alpha_0)_l, l)\}/2)$ where $q_p(a,b) = (1-p)a + pb$, and $f(x) = x$ if $0 \leq x \leq 1$, $f(x) = 0$ if $x < 0$, and $f(x) = 1$ if $x > 1$.

3. If the sampled sharp null maps to a value $DE^0(\alpha_0) \in [\widehat{DE}(\alpha_0)_l, l]$, add $DE^0(\alpha_0)$ to the set $\mathcal{L}$, add the corresponding $p_0$ to $\mathcal{P}$, and if $p_0 \geq \gamma$ then update $l$ to equal $DE^0(\alpha_0)$. Otherwise, do not compute a p-value corresponding to the sampled sharp null and let $p_l = p_l - 1/B_2$.

4. Repeat Steps 2 and 3 $B_2/2 - 1$ times.

The algorithm is modified slightly for $OE(\alpha_0, \alpha_1)$ as it involves all four vectors $\vec{y}(z;\alpha_s)$, $z, s = 0, 1$. Let $\hat{y}(\alpha_s)_l$ and $\hat{y}(\alpha_s)_u$ be the lower and upper limits, respectively, for $\bar{y}(\alpha_s)$ under (3.5). If $p_0 < \gamma$ for $OE(\alpha_0, \alpha_1)_l$, set $l = \widehat{OE}(\alpha_0, \alpha_1)$ and fill in the missingness in $\vec{y}(0;\alpha_0)$ and $\vec{y}(1;\alpha_0)$ with samples from a Bernoulli distribution with mean $f(\{\hat{y}(\alpha_0)_l + \hat{y}(\alpha_1)_u + q_{p_l}(\widehat{OE}(\alpha_0, \alpha_1)_l, l)\}/2)$ where $p_l = 1 - 1/B_2$. A p-value is computed if $OE^0(\alpha_0, \alpha_1) \in [\widehat{OE}(\alpha_0, \alpha_1)_l, l]$ and if not $p_l = p_l - 1/B_2$. If $p_0 \geq \gamma$ for $OE(\alpha_0, \alpha_1)_l$, $l$ is set to equal $OE^0(\alpha_0, \alpha_1)$. The upper endpoint can be approximated using an analogous approach.

Let $t$ be the function from $\mathcal{P}$ to $\mathcal{L}$ that maps each p-value $p_0$ in $\mathcal{P}$ to the null value of $DE^0(\alpha_0)$ in $\mathcal{L}$ which corresponds to the sharp null hypothesis which generated $p_0$. Let $\mathcal{R} = \{\max\{p \in \mathcal{P} : t(p) = l\} : l \in \mathcal{L}\}$. Let $r_1 = \min\{r \in \mathcal{R} : r \geq \gamma\}$ and let $r_2 = \max\{r \in \mathcal{R} : r < \gamma\}$. Let $l_i = t(r_i)$ for $i = 1, 2$. The lower limit of the confidence set $l^*$ is found by local linear interpolation by finding the x-coordinate for the point at which a line drawn from $(l_2, r_2)$ to $(l_1, r_1)$ intersects a horizontal line at $\gamma$, i.e., $l^* = l_2 + (\gamma - r_2)(l_2 - l_1)/(r_2 - r_1)$. The upper limit $u^*$ is found analogously. As $B_2 \to B_1$ and $C_2 \to C_1$, the interval $[l^*, u^*]$ will contain the exact confidence set described in §3.4.1 with probability approaching 1.

The R package `interferenceCI` is available on CRAN (Rigdon, 2015) for computing EIT confidence intervals via this algorithm for the four effects assuming stratified interference when the outcome is binary. The Wald, Chebyshev, and TV intervals are also computed in the package.

## 3.5 Comparisons Via Simulation

A simulation study was carried out to compare the asymptotic, TV, and EIT confidence intervals. The simulation proceeded as follows for fixed values of $\alpha_0$, $\alpha_1$, $DE(\alpha_0)$, $DE(\alpha_1)$, $IE(\alpha_0, \alpha_1)$, $k$, $n_i = n$ for $i = 1, \ldots, k$ such that $N = kn$:

0. Potential outcomes were generated by first fixing the vectors $\vec{y}(z; \alpha_s)$ for $z, s = 0, 1$ to be length $N$ vectors of all 0s. Group membership was assigned by letting elements $n(i-1)+1, \ldots, ni$ of each vector belong to group $i = 1, \ldots, k$. Then, $N(0.5+DE(\alpha_0)/2)$ elements in $\vec{y}(0; \alpha_0)$ were randomly set to equal 1 and $N(0.5 - DE(\alpha_0)/2)$ elements in $\vec{y}(1; \alpha_0)$ were randomly set to equal 1. Then, $N(0.5+DE(\alpha_0)/2-IE(\alpha_0, \alpha_1))$ elements in $\vec{y}(0; \alpha_1)$ were randomly set to equal 1. Finally, $N(0.5+DE(\alpha_0)/2-IE(\alpha_0, \alpha_1)-DE(\alpha_1))$ elements in $\vec{y}(1; \alpha_1)$ were randomly set to equal 1.

1. Observed data were generated by *(i)* randomly assigning $h$ of $k$ groups to strategy $\alpha_1$ and *(ii)* randomly assigning $m_i^s = \alpha_s n$ of $n$ individuals per group to treatment for $s = 0, 1$. Observed outcomes followed based on these treatment assignments and the

potential outcomes from step 0.

2. For each effect, 95% confidence intervals were computed using the observed data generated in step 1.

3. Steps 1-2 were repeated 1000 times.

In the simulation we let $k = n = 10$ or $k = n = 20$ with $h = k/2$, $m_i^0 = 0.3n$ under $\alpha_0$, $m_i^1 = 0.6n$ under $\alpha_1$, $DE(\alpha_0) = 0.95$, $DE(\alpha_1) = 0.3$, and $IE(\alpha_0, \alpha_1) = 0.5$ (such that $TE(\alpha_0, \alpha_1) = 0.8$ and $OE(\alpha_0, \alpha_1) = 0.395$). In the targeted sampling algorithm, $B_2 = C_2 = 100$ such that $B_2/B_1$ and $C_2/C_1$ were less than $10^{-20}$ for all effects. Table 3.1 displays average widths and coverages for Wald, EIT, Chebyshev, and TV. Wald and Chebyshev fail to achieve nominal coverage for $DE(\alpha_0)$ when $k = n = 10$ and Wald additionally fails to cover for $DE(\alpha_0)$ when $k = n = 20$ and for $IE(\alpha_0, \alpha_1)$ and $TE(\alpha_0, \alpha_1)$ when $k = n = 10$. As guaranteed by their respective constructions, EIT and TV achieve nominal coverage for all setups; however, EIT has narrower width than TV in all setups. In fact, EIT is an order of magnitude narrower than TV in three instances: $DE(\alpha_0)$, $TE(\alpha_0, \alpha_1)$, and $OE(\alpha_0, \alpha_1)$ when $k = n = 20$.

## 3.6 Tables and Figures

Table 3.1: Empirical width and coverage [in brackets] of Wald (W), EIT, Chebyshev (C), and TV 95% CIs for simulation study discussed in Section 3.5.

| | $n$ | $k$ | $DE(0.3)$ | $DE(0.6)$ | $IE(0.3, 0.6)$ | $TE(0.3, 0.6)$ | $OE(0.3, 0.6)$ |
|---|---|---|---|---|---|---|---|
| W | 10 | 10 | 0.13 [0.84] | 0.51 [0.96] | 0.39 [0.93] | 0.30 [0.93] | 0.24 [0.94] |
| | 20 | 20 | 0.09 [0.89] | 0.26 [0.96] | 0.21 [0.95] | 0.14 [0.98] | 0.11 [0.97] |
| EIT | 10 | 10 | 0.28 [0.98] | 0.52 [0.98] | 0.47 [0.99] | 0.31 [0.98] | 0.36 [1.00] |
| | 20 | 20 | 0.12 [0.98] | 0.27 [0.98] | 0.24 [0.98] | 0.14 [0.98] | 0.18 [1.00] |
| C | 10 | 10 | 0.22 [0.84] | 1.15 [1.00] | 0.84 [1.00] | 0.54 [1.00] | 0.54 [1.00] |
| | 20 | 20 | 0.15 [0.99] | 0.59 [1.00] | 0.49 [1.00] | 0.32 [1.00] | 0.26 [1.00] |
| TV | 10 | 10 | 1.95 [1.00] | 2.00 [1.00] | 2.00 [1.00] | 2.00 [1.00] | 2.00 [1.00] |
| | 20 | 20 | 1.41 [1.00] | 2.00 [1.00] | 1.86 [1.00] | 1.56 [1.00] | 1.96 [1.00] |

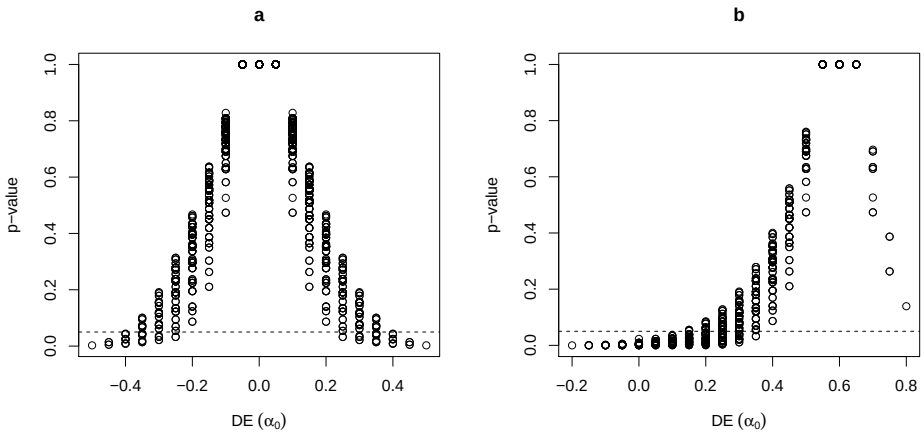


Figure 3.1: Plot of $DE(\alpha_0)$ versus $p(DE(\alpha_0))$ for examples (a) and (b) as outlined in Section 3.4.2.

# CHAPTER 4: BAYESIAN CAUSAL INFERENCE WITH INTERFERENCE

## 4.1 Introduction

Recently, much research has focused on statistical inference in the presence of interference (Hudgens and Halloran, 2008; Tchetgen Tchetgen and VanderWeele, 2012; Aronow, 2012; Bowers et al., 2013). Interference is present when the outcome of one individual is affected by the treatment of any other individual in a randomized experiment or observational study. More generally, interference may occur in any group of humans that communicate or compete with one another. Interference is a violation of the stable unit treatment value assumption (SUTVA) Rubin (1980), a fundamental assumption invoked in many causal evaluations of treatments.

In a two-stage randomized experiment where interference is present, inferential targets have been defined and estimated unbiasedly (Hudgens and Halloran, 2008), and large sample (Liu and Hudgens, 2014) and exact (Tchetgen Tchetgen and VanderWeele, 2012) confidence intervals have been proposed. Statistical methods that do not require adjustments for confounding are appropriate for two-stage randomized experiments, but these experiments are rare. More common are observational studies with clustered data with possible interference in which individuals self-select treatment.

It is possible and even likely in observational studies that the relationship between exposure and outcome is confounded by some other variable. Many strategies have emerged in the causal literature to adjust for such possibilities, two of which are the focus of this paper: inverse probability weighting (IPW) and outcome modeling. As the two dominant schools of statistical thought are frequentist and Bayesian, approaches to evaluating causal exposure-outcome relationships in observational studies with interference may be (1) IPW

frequentist as in Tchetgen Tchetgen and VanderWeele (2012), (2) IPW Bayesian, (3) outcome frequentist, and (4) outcome Bayesian. IPW Bayesian approaches are impossible due to difficulties in constructing a likelihood. An (2010) proposes Bayesian methods using the propensity score for matching and regression. To our knowledge, the literature on Bayesian methods for causal inference in the presence of interference is limited. Toulis and Kao (2013) estimate causal peer influence effects using a frequentist and a Bayesian procedure. Crowley et al. (2014) perform a Bayesian causal analysis on the effects of Haloperidol in paired mice where SUTVA is an untenable assumption.

In this paper, we compare and contrast IPW frequentist, outcome frequentist, and outcome Bayesian approaches to causal inference in an observational study with interference. In Section 2, the motivating example for this research is presented. In Section 3, simulations motivated by the example in Section 2 are conducted without interference. In Section 4, simulations are conducted that allow for interference. In Section 5 the Bayesian methods for observational data with interference are applied to the motivating example. Section 6 concludes with a discussion.

## 4.2 Motivating Example

Researchers studying the preventive effects of bed nets on malaria outcome have observed findings that may suggest interference within communities. In the Wosera area of Papua New Guinea, Hii et al. (2001) found that untreated bed nets have a substantial impact on malaria prevalence in high coverage areas that is greater than can be accounted for by personal protection. In the Nyanza Province in western Kenya, Hawley et al. (2003) found a similar strong community effect of treated bed nets, hypothesizing that these effects were due to reduced longevity of mosquito populations forced to expend extra energy in search of human blood.

The 2007 Demographic and Health Survey in the Democratic Republic of the Congo, henceforth DHS, was a nationally representative survey designed to provide information on fertility, mortality, sexually transmitted infections, mosquito net use, and malaria outcome,

among other health issues. In total, nine thousand households were surveyed. Malaria outcome was determined by real-time PCR assays on genomic DNA from dried blood spots. Data exist for 8,844 individuals in 300 communities ranging in size from 4 to 52 individuals with a mean of 29.48. In the DHS, 20% of individuals self-reported using a bed net the previous night. The proportion of individuals within each group who reported using a bed net the previous night had 1st, 2nd, and 3rd quartiles equal to 4%, 14%, and 32%, respectively. Messina et al. (2011) modeled malaria status in these data as a function of group and individual covariates in a multilevel logistic regression. Although individual bed net usage was significant when entered alone in the model, it was no longer significant when group level bed net coverage was included in the model.

Table 4.1 suggests that interference may be present in the DHS by displaying malaria incidence in those with and without bed nets while grouping the 300 communities by septiles of community bednet usage. Those individuals without bed nets in the first quartile of group coverage were 4% more likely to contract malaria as those in the fourth quartile of group coverage. In Section 5, a re-analysis of these data are conducted that allows for interference.

## 4.3 Observational Study Inference Without Interference

Consider an observational study without interference with binary outcome $Y$ (e.g. malaria), binary exposure $Z$ (e.g., bed net use), and covariates $X$. For individual $j = 1, \ldots, n$, let $Y_j(z)$ be the outcome at exposure level $z = 0, 1$. Let $X_j$ be an observed covariate, and $Z_j$ be the observed exposure. The observed outcome is $Y_j = Z_j Y_j(1) + (1 - Z_j)Y_j(0)$. Assuming that $(Y(0), Y(1), Z, X)$ are random variables, the target of inference is the population average causal effect, $\tau = E[Y(0)] - E[Y(1)]$. As the outcome is binary, $\tau \in [-1, 1]$. If $Y$ is malaria outcome and $Z$ is bed net use, values of $\tau$ close to 1 indicate significant preventive effects of bed net use, values close to 0 indicate no effect, and values close to -1 indicate harm caused by bed net use.

In a randomized trial, inference on $\tau$ is straightforward using well known methods. The

estimator

$$\hat{\tau} = \frac{1}{n}\sum_{j=1}^{n}\frac{1\{Z_j = 0\}Y_j}{\Pr[Z_j = 0]} - \frac{1}{n}\sum_{j=1}^{n}\frac{1\{Z_j = 1\}Y_j}{\Pr[Z_j = 1]} \tag{4.1}$$

is unbiased because $Y(0), Y(1) \perp Z$, i.e., treatment assignment is independent of potential outcomes. The quantity $\Pr[Z_j = z] \equiv n_z/n$ where $n_z = \sum_{j=1}^{n} 1\{Z_j = z\}$ is known in advance so that $\hat{\tau} = \hat{p}_0 - \hat{p}_1$ where $p_z = \sum_{j=1}^{n} 1\{Z_j = z\}Y_j/n_z$ for $z = 0, 1$, is the familiar difference in proportions.

In a randomized trial, a commonly used interval estimator for $\tau$ is the large sample Wald confidence interval

$$\hat{\tau} \pm z_{1-\alpha/2}\sqrt{\hat{p}_0(1-\hat{p}_0)/n_0 + \hat{p}_1(1-\hat{p}_1)/n_1} \tag{4.2}$$

where $z_c$ is the $c^{th}$ quantile of a standard normal distribution. The interval in (4.2) will contain $\tau$ with probability $1 - \alpha$ as $n \to \infty$.

### 4.3.1 Inverse Probability Weighted Estimation

Inference on $\tau$ in an observational study is more difficult as it is unlikely that treatment assignment is independent of potential outcomes and $\Pr[Z_j = z]$ is not known in advance. The naive estimator $\hat{\tau}$ will be subject to confounding, in which the same characteristics that lead an individual to be exposed to a treatment may also be associated with the potential outcome. When all of the confounders $X$ are measured, it may be reasonable to assume

$$Y(0), Y(1) \perp Z|X \tag{4.3}$$

i.e., strong ignorability (Rosenbaum and Rubin, 1983). Suppose further that the relationship between $Z$ and $X$ is represented by the propensity score $e(X) = \Pr[Z = 1|X]$. In such a case, the estimator

$$\hat{\tau}^{IPW} = \frac{1}{n}\sum_{j=1}^{n}\frac{1\{Z_j = 0\}Y_j}{1 - e_j(X)} - \frac{1}{n}\sum_{j=1}^{n}\frac{1\{Z_j = 1\}Y_j}{e_j(X)} \tag{4.4}$$

is unbiased. In practice, $e(X)$ is rarely known and must be estimated using logistic regression or some other modeling strategy, e.g.

$$\hat{\tau}_e^{IPW} = \frac{1}{n}\sum_{j=1}^{n}\frac{1\{Z_j = 0\}Y_j}{1-\hat{e}_j(X)} - \frac{1}{n}\sum_{j=1}^{n}\frac{1\{Z_j = 1\}Y_j}{\hat{e}_j(X)} \tag{4.5}$$

When the propensity score is known, $\hat{\tau}^{IPW}$ is consistent for $\tau$ and asymptotically Normal, and when the propensity score is estimated $\hat{\tau}_e^{IPW}$ is consistent for $\tau$ and asymptotically Normal (Lunceford and Davidian, 2004).

### 4.3.2 Outcome Modeling

An alternative to IPW is outcome modeling. It follows from (4.3) that for $z = 0, 1$

$$\begin{aligned} E_X[E_{Y|X}(Y|Z = z, X)] &= E_X[E_{Y|X}(Y(z)|Z = z, X)] \\ &= E_X[E_{Y|X}(Y(z)|X)] \\ &= E[Y(z)] \end{aligned} \tag{4.6}$$

Thus, $\tau$ can be directly estimated using outcome (regression) models for $E[Y|Z, X]$. The parameters $\beta$ in the outcome model are estimated using observed data. The model predicted potential outcomes are used to estimate $\tau$ as follows

$$\hat{\tau}^o = \frac{1}{n}\sum_{j=1}^{n}[E\{Y_j(0)|Z, X, \hat{\beta}\} - E\{Y_j(1)|Z, X, \hat{\beta}\}] \tag{4.7}$$

The large sample distribution and variance of $\hat{\tau}^o$ can be found using M-estimation theory (Stefanski and Boos, 2002).

### 4.3.3 A Bayesian Approach

Inference on $\tau$ can be carried out using a Bayesian approach. Let $\theta$ be a vector of parameters of interest, and let $Y = Y(0), Y(1)$ be the potential outcomes. Partition $Y = (Y^{obs}, Y^{mis})$

where $Y^{obs}$ is the part of $Y$ sampled by $Z$ in the study, and $Y^{mis}$ are the unobserved potential outcomes. Define $X = (X^{obs}, X^{mis})$ similarly, but for now assume that $X$ is fully observed so that $X = X^{obs}$. Inference on $\tau$ is carried out using the posterior distribution $f(\tau|Y^{obs}, X^{obs}, Z)$. Under (4.3), sampling from the posterior distribution can be accomplished through specifying the conditional distribution of $Y$ given $X$, $Z$, and $\theta$, $f(Y|X^{obs}, Z, \theta)$, and a prior distribution on $\theta$, $f(\theta)$ (Rubin, 1978, Section 4). The following Gibbs sampler can be used to sample from the posterior distribution:

1. Sample initial values for $\theta$ from the prior $f(\theta)$.
2. Sample values for $Y^{mis}$ conditional on $\theta$ from $f(Y^{mis}|Y^{obs}, X^{obs}, Z, \theta)$. Causal effects that are functions of $Y$, $\tau(Y)$, can be computed.
3. Sample values for $\theta$ from $f(\theta|Y^{obs}, Y^{mis}, X^{obs}, Z)$, i.e., conditional on the sampled $Y^{mis}$ in step 2. Causal effects that are functions of $\theta$, $\tau(\theta)$, can be computed.
4. Iterate until convergence.

If inference concerns the sample average causal effect, then $f(\tau(Y)|Y^{obs}, X^{obs}, Z)$ is the target of inference. If inferences are being drawn about the population average causal effect (as in this paper), then $f(\tau(\theta)|Y^{obs}, X^{obs}, Z)$ is the target of inference. The posterior mean and credible interval may serve as point and interval estimates, respectively, of the causal effect $\tau$. In contrast to the outcome modeling strategy, the Bayesian strategy can deal with missingness in $X$ in a simple and intuitive manner.

### 4.3.4 A Simulation Study

The naive, IPW, outcome, and Bayesian point and interval estimators were evaluated via a simulation study motivated by the DHS. For individual $j = 1, \ldots, n$, let $Y_j(z)$ be the potential malaria outcome at exposure level $z = 0, 1$, let $X_j$ be the binary covariate proximity to an urban space, and let $Z_j$ be the observed bed net use. The simulation proceeded as follows:

The simulation proceeded as follows:

1. For individual $j = 1, \ldots, n$, $X_j$ was randomly sampled from $f(X)$
2. For each individual $Z_j$ was self-selected from a Bernoulli distribution with mean $\mathcal{L}^{-1}(\gamma_0 + \gamma_1 X_j)$ in which $X_j$ was observed in step 1
3. For each individual the malaria outcome $Y_j$ was sampled from a Bernoulli distribution with mean $\mathcal{L}^{-1}(\beta_0 + \beta_1 Z_j + \beta_2 X_j)$
4. Point estimates and 95% confidence (or credible) intervals for $\tau$ were computed using the naive, inverse probability weighted, outcome regression, and Bayesian approaches. In the Bayesian approach, the burn-in was 1100, no thinning was used, and 10000 samples from the posterior were taken.
5. Steps 1-4 were repeated 1000 times.

Consequently, the true causal effect was equal to

$$\tau = \int_X \{\mathcal{L}^{-1}(\beta_0 + \beta_2 X) - \mathcal{L}^{-1}(\beta_0 + \beta_1 + \beta_2 X)\} f(X) dX \tag{4.8}$$

where here and in the sequel $\mathcal{L}(x)$ is shorthand for $\text{logit}(x) = \log\{x/(1-x)\}$.

In the Bayesian approach in the simulation above, non-informative priors were used on all model parameters, i.e., $\beta_0, \beta_1, \beta_2 \overset{iid}{\sim} \mathcal{N}(0, 4)$ such that 95% of the prior probability for all odds ratios was in [0.02, 50.4]. Additionally, the correct outcome model $\Pr[Y_j(z) = 1|X_j]$ was used in the regression and Bayesian approaches and the correct propensity score model $\Pr[Z_j = 1|X_j]$ was used in the IPW estimator.

Table 4.2 displays simulation results in which the true parameters were estimated using the DHS data such that $\beta_0 = -0.49$, $\beta_1 = -0.39$, $\beta_2 = -0.63$, $\gamma_0 = -1.61$, $\gamma_1 = 0.47$, and where $X_j$ were randomly sampled from a Bernoulli distribution with mean 0.45 implying that $\tau = 0.077$. The results in Table 4.2 confirm that IPW and Bayes both adjust for the confounding present in the naive estimator. As pointed out in Lunceford and Davidian

(2004), the IPW estimator with estimated propensity score has narrower width than the IPW estimator with known propensity score. The frequentist outcome, IPW, and Bayesian approaches all have similar widths for $n = 50$, 200, and 1000; however, only the Bayesian approach has nominal 95% coverage for $n = 50$. In the next section, a simulation study is conducted mirroring the one in this section with the additional feature of interference.

## 4.4 Observational Study Inference with Interference

In this section, methods for inference with interference in observational studies are reviewed and a new approach is proposed. Let $Y_{ij}(z_i)$ be the binary outcome (e.g., malaria) for individual $j = 1, \ldots, n_i$ in group $i = 1, \ldots, k$ when group $i$ has exposure $z_i$ (e.g., bed net use). Let $z_i \in \mathcal{Z}(n_i)$ where $Z(n_i)$ contains the $2^{n_i}$ realizations of $Z_i$, e.g., $\mathcal{Z}(2) = \{(0,0),(0,1),(1,0),(1,1)\}$. The vector of exposures can be partitioned as $z_i = (z_{ij}, z_{i(j)})$ where $z_{ij}$ is the treatment assignment for individual $j$ and $z_{i(j)}$ is the treatment assignment for all individuals except individual $j$. In this formulation, both $Y_{ij}(z_i)$ and $Z_i$ are random variables. Throughout, partial interference (Sobel, 2006) is assumed such that interference can only occur within groups. Thus, each individual's potential outcomes are a set of random variables $\{Y_{ij}(z_i) : z_i \in \mathcal{Z}(n_i)\}$. Suppose that each individual independently selects treatment with probability $\alpha$. In this scenario, each individual's set of potential outcomes can be summarized by weighted averages. Let $\pi_\alpha(\omega|z) = \Pr_\alpha(z_{i(j)} = \omega | z_{ij} = z)$. The average potential outcome for individual $j$ in group $i$ given $z_{ij} = z$ equals

$$\begin{aligned}\bar{Y}_{ij}(z;\alpha) &= \sum_{\omega \in \mathcal{Z}(n_i-1)} \pi_\alpha(\omega|z) Y_{ij}(z_{ij} = z; z_{i(j)} = \omega) \\ &= \sum_{\omega \in \mathcal{Z}(n_i-1)} \alpha^{|\omega|}(1-\alpha)^{n_i-1-|\omega|} Y_{ij}(z_{ij} = z; z_{i(j)} = \omega)\end{aligned}$$

in which $z_{i(j)} \in \mathcal{Z}(n_i - 1)$ is the treatment vector of all individuals in group $i$ except $j$, and $|\omega|$ denotes the sum of $\omega$. The group $i$ average potential outcome on treatment $z = 0, 1$ under treatment strategy $\alpha$ equals $\bar{Y}_i(z;\alpha) = n_i^{-1} \sum_{j=1}^{n_i} \bar{Y}_{ij}(z;\alpha)$. Finally, let $F_{z,\alpha}$ be the

distribution function of $\bar{Y}_i(z;\alpha)$, and define

$$\mu_{z,\alpha} \equiv E[\bar{Y}_i(z;\alpha)] = \int_0^1 y dF_{z,\alpha}(y)$$

Let $\pi_\alpha(x) = \Pr_\alpha(z_i = x)$. Assuming that each individual independently selects treatment with probability $\alpha$, the average potential outcome for individual $j$ in group $i$ under $\alpha$ is equal to

$$\begin{aligned}
\bar{Y}_{ij}(\alpha) &= \sum_{\omega \in \mathcal{Z}(n_i)} \pi_\alpha(\omega) Y_{ij}(z_i = \omega) \\
&= \sum_{\omega \in \mathcal{Z}(n_i)} \alpha^{|\omega|}(1-\alpha)^{n_i - |\omega|} Y_{ij}(z_i = \omega) \\
&= \sum_{z=0,1} \sum_{\omega' \in \mathcal{Z}(n_i - 1)} \alpha^z (1-\alpha)^{1-z} \alpha^{|\omega'|}(1-\alpha)^{n_i - 1 - |\omega'|} Y_{ij}(z_{ij} = z; z_{i(j)} = \omega') \\
&= (1-\alpha)\bar{Y}_{ij}(0;\alpha) + \alpha \bar{Y}_{ij}(1;\alpha) \qquad (4.9)
\end{aligned}$$

The group $i$ average potential outcome under treatment strategy $\alpha$ is equal to $\bar{Y}_i(\alpha) = n_i^{-1} \sum_{j=1}^{n_i} \bar{Y}_{ij}(\alpha)$. Finally, let $\mu_\alpha \equiv E[\bar{Y}_i(\alpha)] = (1-\alpha)\mu_{0,\alpha} + \alpha \mu_{1,\alpha}$.

Extending Hudgens and Halloran (2008) to the superpopulation setting, the inferential targets are

$$\begin{aligned}
DE(\alpha) &= \mu_{0,\alpha} - \mu_{1,\alpha} \\
IE(\alpha, \alpha') &= \mu_{0,\alpha} - \mu_{0,\alpha'} \\
TE(\alpha, \alpha') &= \mu_{0,\alpha} - \mu_{1,\alpha'} \\
OE(\alpha, \alpha') &= \mu_\alpha - \mu_{\alpha'} \qquad (4.10)
\end{aligned}$$

The observed data reveal for group $i = 1, \ldots, k$ the treatment vector $Z_i = (Z_{i1}, \ldots, Z_{in_i})$, covariates $X_i$, and outcome vector $Y_i = (Y_{i1}, \ldots, Y_{in_i})$.

### 4.4.1 Inverse Probability Weighted Estimation

Tchetgen Tchetgen and VanderWeele (2012) showed that

$$\hat{Y}_i^{IPW}(z;\alpha) = \frac{1}{n_i}\sum_{j=1}^{n_i} \frac{\pi_\alpha(Z_{i(j)}|Z_{ij})1\{Z_{ij}=z\}Y_{ij}}{\Pr[Z_i|X_i]} \tag{4.11}$$

and

$$\hat{Y}_i^{IPW}(\alpha) = \frac{1}{n_i}\sum_{j=1}^{n_i} \frac{\pi_\alpha(Z_i)Y_{ij}}{\Pr[Z_i|X_i]} \tag{4.12}$$

are unbiased for $\bar{Y}_i(z;\alpha)$ and $\bar{Y}_i(\alpha)$, respectively, in a finite population setup assuming (4.3) when the propensity score $\Pr[Z_i|X_i]$ is known. The estimators (4.11) and (4.12) are also unbiased in the superpopulation setting, as in the following propositions:

**Proposition 4.1.** $E[\hat{Y}_i^{IPW}(z;\alpha)] = \mu_{z,\alpha}$

**Proposition 4.2.** $E[\hat{Y}_i^{IPW}(\alpha)] = \mu_\alpha$

Proofs of Proposition 4.1 and 4.2 are given in the Appendix. In observational studies, $\Pr[Z_i = z|X_i]$ is rarely known and must be estimated, e.g., by a regression model. The product of individual level model predicted probabilities serves as the estimator $\widehat{\Pr}[Z_i|X_i]$. When the propensity score is known or estimated correctly, Perez-Heydrich et al. (2014, Web appendix) use M-estimation theory to show that the IPW estimators in (4.4) are consistent for their target parameters and asymptotically normal. Additionally, they present sandwich variance estimators to be used in confidence intervals.

### 4.4.2 Outcome Modeling

As noted previously, under partial interference, individual $j$ in group $i$ has $2^{n_i}$ potential outcomes. As $n_i$ increases, the computational difficulties associated with this problem mount considerably. In many settings, it may be reasonable to consider functional assumptions about interference

$$Y_{ij}(z_{ij}=z, z_{i(j)}=\omega) = Y_{ij}(z_{ij}=z, z_{i(j)}=\omega') \;\forall\; \omega,\omega' \in \mathcal{Z}(n_i-1) \text{ s.t. } f(\omega)=f(\omega') \tag{4.13}$$

Under (4.13), an individual's potential outcome on treatment $z = 0, 1$ is the same when some function of the other $n_i - 1$ treatment assignments maps to the same value. At one extreme, when $f(x) = x$, each individual has $2^{n_i}$ potential outcomes, and at the other extreme when $f(x) = 0$, each individual only has two potential outcomes (no interference). One reasonable function is $f(x) = 1\{|x| > 0\}$, where $|x|$ denotes the sum of the elements in $x$. Under this threshold function, each individual has four potential outcomes, $\{y_{ij}(z_{ij} = z, 1\{|z_{i(j)}| > 0\} = c)\}$ for $z, c = 0, 1$. Another reasonable assumption is stratified interference, or the function $f(x) = |x|$. Under this assumption, each individual has $2n_i$ potential outcomes. The purpose of the functional assumption (4.13) is to simplify the problem, mapping the $2^{n_i}$ group vector assignments to the real line. Under (4.13), for $z = 0, 1$

$$\bar{Y}_{ij}(z; \alpha) = \sum_{c \in \mathcal{C}} \left\{ \sum_{\omega: f(\omega) = c} \pi_\alpha(z_{i(j)} = \omega | z_{ij} = z) \right\} Y_{ij}(z_{ij} = z; f(z_{i(j)}) = c)$$

A statistical model for $E[Y_{ij}(z_{ij} = z; f(z_{i(j)}) = c)]$ can be used for inference on the causal effects (4.10), e.g.,

$$\mathcal{L}(\Pr[Y_{ij}(z_{ij} = z; f(z_{i(j)}) = c) = 1]) = \beta_0 + \beta_1 z + \beta_2 c + \eta X_i \tag{4.14}$$

in which $\Pr[Y_{ij}(z_{ij} = z; f(z_{i(j)}) = c) = 1] = E[Y_{ij}(z_{ij} = z; f(z_{i(j)}) = c)]$ as $Y$ is binary. The parameter estimates for model (4.14), the resulting causal effect estimators, and the corresponding large sample distributions can be derived also using M-estimation theory.

### 4.4.3 A Bayesian Approach

A Bayesian approach to evaluating causal exposure-outcome relationships in observational studies with interference is outlined in this section. Although similar to the outcome modeling approach, the Bayesian approach has the key advantage of accommodating missing covariate data in a straightforward manner, an important feature given that the DHS contain missing covariate data. Using model (4.14) for the potential outcomes, let $\theta = (\beta_0, \beta_1, \beta_2, \eta)$. Inference on the causal effects is carried out by sampling from their posterior distributions using an analogous Gibbs sampler to the one outlined in Section 2. In this Gibbs sampler, each

individual will have greater than or equal to two potential outcomes, so that $Y^{mis}$ will be of length greater than or equal to one for each individual. The causal effects (4.10) are functions of $\theta$ and can be directly computed in step 3 of the Gibbs sampler.

### 4.4.4 A Simulation Study

To study the proposed methods, the naive, IPW, outcome, and Bayesian point and interval estimators were compared in a simulation study motivated by the DHS that assumed stratified interference. As will become clear in the simulation below, the true causal effects were functions of $\mu_{z,\alpha} =$

$$E_{X,n_i}\frac{1}{n_i}\sum_{j=1}^{n_i}\sum_{c=0}^{n_i-1}\binom{n_i-1}{c}\alpha^c(1-\alpha)^{n_i-1-c}\mathcal{L}^{-1}\left(\beta_0+\beta_1 z+\beta_2\frac{z+c}{n_i}+\beta_3 X_{ij}+\beta_4 z\frac{z+c}{n_i}\right) \tag{4.15}$$

and $\mu_\alpha = (1-\alpha)\mu_{0,\alpha}+\alpha\mu_{1,\alpha}$. The derivation of (4.15) is given in the appendix.

The simulation proceeded as follows:

1. For group $i = 1, \ldots, k$, $n_i$ was randomly sampled from $f(n_i)$
2. For individual $j$ in group $i$, a $X_{ij}$ was randomly sampled from $f(X)$
3. For individual $j$ in group $i$, bed net status $Z_{ij}$ was self-selected using the following model

$$Z_{ij} \sim \text{Bernoulli}(\mathcal{L}^{-1}(\gamma_0+\gamma_1 X_{ij})) \tag{4.16}$$

4. Given $Z_i = (Z_{ij}, Z_{i(j)})$ from step 3, an observed malaria outcome for individual $j$ in group $i$ was generated using the model

$$Y_{ij}(z_{ij}=z, z_{i(j)}=\omega) \sim \text{Bernoulli}(\mathcal{L}^{-1}(\beta_0+\beta_1 z+\beta_2(|\omega|+z)/n_i+\beta_3 X_{ij}+\beta_4 z(|\omega|+z)/n_i)) \tag{4.17}$$

5. To mimic a two-stage randomized experiment, for groups where $\sum_j Z_{ij}/n_i \in [\alpha_s - w, \alpha_s + w]$, $S_i$ was set to equal $s = 0, 1$. Groups who did not meet these criteria

were excluded and the estimators of Hudgens and Halloran (2008) and Wald-type 95% confidence intervals of Liu and Hudgens (2014) were computed as the naive estimators.

6. IPW estimators and 95% confidence intervals were computed wherein the propensity scores were treated as known.

7. IPW estimators and 95% confidence intervals were computed wherein the propensity scores were estimated using the correct model (4.16).

8. Outcome frequentist estimators and 95% confidence intervals were computed using the correct model (4.17).

9. Bayesian posterior means and 95% credible intervals were computed using the correct model (4.17) and the priors $\beta_m \sim \text{Normal}(0, 4)$ for $m = 0, \ldots, 3$ with a burn-in of 1100, a thinning interval of 3, and 1000 samples for the posterior distribution.

10. Steps 1-9 were repeated 1000 times.

The simulations with interference were also motivated by the DHS such that for individual $j$ in group $i$, $Y_{ij}(z_i)$ was the binary potential malaria outcome, $Z_{ij}$ was the binary exposure of bed net use, $\sum_j Z_{ij}/n_i$ was the proportion of bed net use in group $i$, and $X_{ij}$ was the binary confounder of proximity to an urban space. The distributions of $n_i$ and $X$ were estimated from the DHS such that $n_i$ ranged from 4-52 with a mean of 29.48 and such that $\Pr[X = 1] = 0.45$. In computing the naive estimator, $w = 0.05$. The parameters in models (4.16) and (4.17) were estimated using the DHS such that $\gamma_0 = -1.61$, $\gamma_1 = 0.47$, $\beta_0 = -0.47$, $\beta_1 = -0.40$, $\beta_2 = -0.13$, $\beta_3 = -0.63$, $\beta_4 = -0.06$. Additional simulation inputs are summarized in Figure 4.1.

Simulation results are given in Figure 4.2 for $\alpha = 0, 0.1, \ldots, 1$. The IPW estimators can be biased, can have ESE/ASE ratios much larger than 1, can fail to achieve nominal coverage, and can have much larger widths than outcome models whereas the outcome models are unbiased, have ESE/ASE ratios near 1, achieve nominal coverage, and have sensible widths. The overall performance of the IPW estimators was noticeably worse for $\alpha$ values in low probability areas

of the observed range of group coverage, i.e., $\alpha \leq 0.1$ or $\alpha \geq 0.4$. For example, IPW estimators for $\alpha = 0.2$ had good operating characteristics whereas IPW estimators for $\alpha = 0.8$ had poor operating characteristics.

## 4.5 Analysis of the DHS

The Bayesian method discussed above was applied to the DHS. Based on expert opinion, the confounders age, sex, urban, time to water source, roof materials, density of agriculture in 10 km, rain, and air temperature in celsius were included in the outcome model for malaria in addition to bed net at the individual and group level. The variables malaria, bed net, age, sex, and urban were fully observed whereas time to water source, roof materials, agricultural density, rain, and air temperature had varying degrees of missingness less than 2.3%.

In the first Bayesian model, only fully observed variables were entered into the outcome model

$$\begin{aligned}\mathcal{L}(\Pr[\text{malaria}_{ij} = 1]) = b_0 + b_1\text{bednet}_{ij} + b_2\text{prop}_i + b_3\text{age}_{ij} + b_4\text{sex}_{ij} \\ + b_5\text{urban}_{ij} + b_6\text{bednet}_{ij}\text{prop}_i\end{aligned} \tag{4.18}$$

and Normal$(0, 4)$ priors were specified for all model parameters. A burn-in of 1100, a thinning interval of 3, and 1000 samples for the posterior distribution were used. Results are summarized in Table 4.3 for $\alpha \in \{0.2, 0.5, 0.8\}$.

In the second Bayesian model, all of the variables were entered into the outcome model

$$\begin{aligned}\mathcal{L}(\Pr[\text{malaria}_{ij} = 1]) = b_0 + b_1\text{bednet}_{ij} + b_2\text{prop}_i + b_3\text{age}_{ij} + b_4\text{sex}_{ij} \\ + b_5\text{urban}_{ij} + b_6\text{time2H2O}_{ij} + b_7\text{roof}_{ij} \\ + b_8\text{agric10km}_{ij} + b_9\text{rain}_{ij} + b_{10}\text{temp}_i + b_{11}\text{bednet}_{ij}\text{prop}_i\end{aligned} \tag{4.19}$$

and the variables with missingness were modeled in lower level hierarchies: time2H2O as a function of urban, roof as a function of urban, agric10km as a function of urban, rain as a function of temp and temp with only an intercept. Again Normal$(0, 4)$ priors were specified for all model parameters and the hierarchical models for the missing variables were normally distributed with common variance $\sigma^2$ that had a Uniform$(0, 100)$ prior. A burn-in of 1100, a thinning interval of 3, and 1000 samples for the posterior distribution were used. Results are summarized in Table 4.3 for $\alpha \in \{0.2, 0.5, 0.8\}$.

Model 4.19 may be more plausible than model 4.18 as it handles missing data in a reasonable manner rather than discarding any observations with missing data as in model 4.18. Complete case analyses such as model 4.18 may lead to bias (Little and Rubin, 1989), so the inferences from model 4.19 will be discussed. When treating 20% of the population with a bed net, the direct effect of bed net on malaria is about a 7% reduction. Treating 50% and 80% of the population reduces this direct effect to about a 6% reduction in malaria.

The spillover effect is most pronounced when $\alpha_0 = 0.2$ and $\alpha_1 = 0.8$ such that an individual not using a bed net in a group with 20% coverage is about 3% more likely to contract malaria relative to an individual not using a bed net in a group with 80% coverage. This indirect effect narrowly reaches statistical significance as the lower limit to the credible interval is greater than zero.

The total effect and overall effects also reach statistical significance at $\alpha_0 = 0.2$ and $\alpha_1 = 0.8$. An individual not using a bed net in a group with 20% coverage is about 8% more likely to contract malaria than an individual using a bed net in a group with 80% coverage. Any individual in a group with 20% coverage is about 6% more likely to contract malaria than an individual in a group with 80% coverage.

## 4.6 Discussion

The finding of a spillover effect that is statistically significant is in line with the Messina et al. (2011) finding that community level bed net use is a significant predictor of malaria

outcome whereas individual level use is not. The magnitude of the spillover effect was about 1/2 as large as the direct effect of bed net use. A spillover effect beyond the direct effect of bed net use may be in line with findings of Hii et al. (2001) and Hawley et al. (2003). This important public health finding lends more support to the argument that sleeping under bed nets in the DRC should we a widespread practice if malaria prevention is the goal.

One key difference between our methods and the methods in Messina et al. (2011) is that our approach only used one bed net variable (whether or not an individual slept under treated net last night) to define individual and community level bed net use whereas Messina et al. (2011) used four (number of household bed nets, whether or not household has bed net, whether or not an individual slept under treated net last night, and whether or not an individual slept under an untreated net last night). In fact in the DHS the variable whether or not household has bed net is a function of the number of household bed nets.

As individuals self-selected bed nets in the DHS, Bayesian outcome models for malaria risk were used to adjust for the potential confounding in the DHS. The Bayesian outcome model may be preferred to the frequentist outcome model as it does not rely upon large sample theory and can handle missing covariate data in an intuitive manner. The Bayesian outcome model may be preferred to IPW estimation due to better operating characteristics in this setting as shown in the simulation study in section 4.4.4. Other papers confirm the point that IPW estimators have the potential for instability when weights are small enough (Kang and Schafer, 2007; Freedman and Berk, 2008).

A future direction of this research is to carry out similar Bayesian causal analyses in the DHS using the other three bed net variables rather than whether or not household has a bed net to further explore the relationship between bed net use and malaria outcome at the individual and community level. The reliability and validity of self-report of whether or not an individual slept under a bed net last night as a proxy of bed net use could also be investigated.

## 4.7 Tables and Figures

Table 4.1: Estimated malaria outcome by bed net status in 300 communities in 2007 DHS stratified by quartile of community bed net use

| | No bed net | | | Bed net | | | Total | | |
|---|---|---|---|---|---|---|---|---|---|
| Quartile | Malaria | Total | Prop | Malaria | Total | Prop | Malaria | Total | Prop |
| 0-4 | 631 | 1817 | 0.35 | 107 | 409 | 0.26 | 738 | 2226 | 0.33 |
| 4-14 | 531 | 1718 | 0.31 | 118 | 506 | 0.23 | 649 | 2224 | 0.29 |
| 14-32 | 567 | 1769 | 0.32 | 79 | 424 | 0.19 | 646 | 2193 | 0.29 |
| 32-100 | 545 | 1762 | 0.31 | 107 | 439 | 0.24 | 652 | 2201 | 0.30 |

Table 4.2: Empirical bias and variance for point estimators of $\tau$, and width and coverage for interval estimators of $\tau$ where $n$ is sample size

| | | Point estimator | | | Interval estimator | |
|---|---|---|---|---|---|---|
| Method | $n$ | Bias | ESE | ASE | Coverage | Width |
| Naive | 50 | 0.015 | 0.15 | 0.15 | 0.88 | 0.57 |
| | 200 | 0.017 | 0.081 | 0.075 | 0.91 | 0.29 |
| | 1000 | 0.016 | 0.035 | 0.034 | 0.91 | 0.13 |
| IPW oracle | 50 | 0.0013 | 0.18 | 0.17 | 0.90 | 0.68 |
| | 200 | 0.0029 | 0.096 | 0.090 | 0.92 | 0.35 |
| | 1000 | 0.0020 | 0.041 | 0.041 | 0.95 | 0.16 |
| IPW | 50 | 0.0015 | 0.16 | 0.14 | 0.85 | 0.55 |
| | 200 | 0.0044 | 0.083 | 0.077 | 0.92 | 0.30 |
| | 1000 | 0.0018 | 0.036 | 0.035 | 0.94 | 0.14 |
| Outcome | 50 | 0.0013 | 0.15 | 0.15 | 0.88 | 0.58 |
| | 200 | 0.0033 | 0.082 | 0.076 | 0.92 | 0.30 |
| | 1000 | 0.0022 | 0.036 | 0.035 | 0.94 | 0.14 |
| Bayes | 50 | -0.013 | 0.12 | 0.13 | 0.97 | 0.52 |
| | 200 | 0.00052 | 0.078 | 0.074 | 0.93 | 0.29 |
| | 1000 | 0.0017 | 0.035 | 0.034 | 0.94 | 0.13 |

Table 4.3: Results of Bayesian models for DHS data

| | Model 4.18 | | | Model 4.19 | | |
|---|---|---|---|---|---|---|
| Effect | Estimate | 95% Credible Interval | | Estimate | 95% Credible Interval | |
| $DE(0.2)$ | 0.0670 | 0.0440 | 0.090 | 0.063 | 0.0380 | 0.084 |
| $DE(0.5)$ | 0.0620 | 0.0210 | 0.100 | 0.060 | 0.0190 | 0.097 |
| $DE(0.8)$ | 0.057 | -0.0180 | 0.130 | 0.054 | -0.01500 | 0.120 |
| $IE(0.2, 0.5)$ | 0.0082 | -0.0068 | 0.024 | 0.013 | -0.0015 | 0.028 |
| $IE(0.2, 0.8)$ | 0.016 | -0.0150 | 0.044 | 0.028 | 0.00065 | 0.055 |
| $TE(0.2, 0.5)$ | 0.0700 | 0.0300 | 0.110 | 0.073 | 0.0350 | 0.110 |
| $TE(0.2, 0.8)$ | 0.073 | 0.0038 | 0.140 | 0.083 | 0.01800 | 0.140 |
| $OE(0.2, 0.5)$ | 0.0260 | 0.0061 | 0.044 | 0.031 | 0.0140 | 0.047 |
| $OE(0.2, 0.8)$ | 0.048 | -0.0030 | 0.098 | 0.059 | 0.00730 | 0.100 |

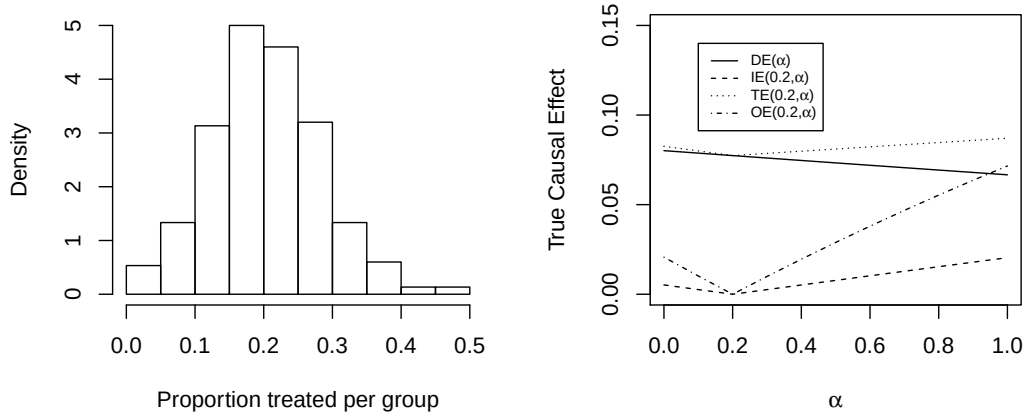


Figure 4.1: Summary of inputs in simulation study in Section 4.4.4

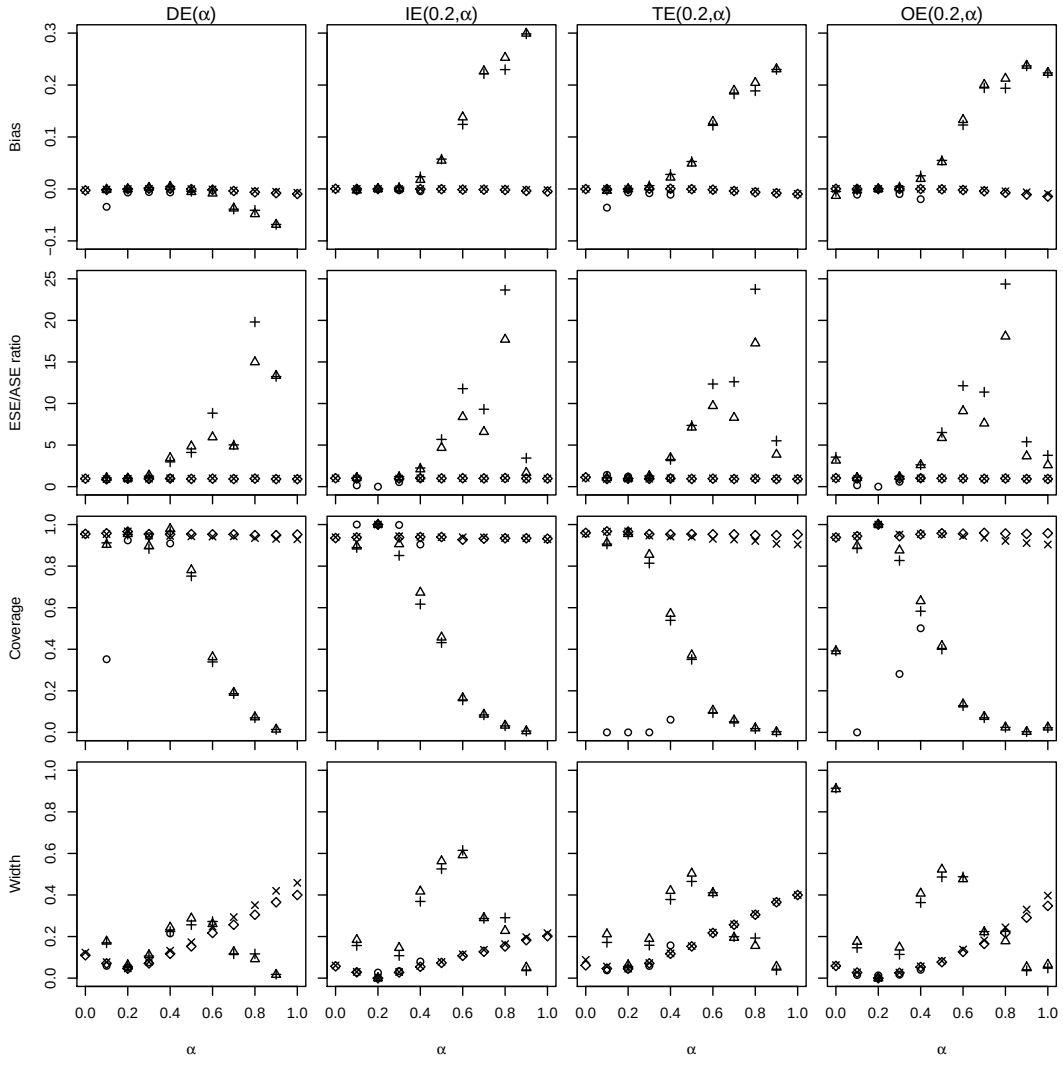


Figure 4.2: Summary of operating characteristics in simulation study in Section 4.4.4 where circle=Naive, triangle=IPW, plus=IPW estimated PS, x=outcome, and diamond=Bayes

## APPENDIX A: TECHNICAL DETAILS FOR CHAPTER 2

*Proof of Proposition 2.1*

Let $L^D = L^0 + L^1$, $U^D = U^0 + U^1$, $A^1 = A^1(Z, \delta)$, and $A^0 = A^0(Z, \delta)$. Recall that $n\tau = \sum \delta_j$. After observing the data, $A^1 \in \mathcal{A}^1$, where $\mathcal{A}^1 = \{-\sum Z_j Y_j, ..., \sum Z_j(1 - Y_j)\}$. Therefore:

$$
\begin{aligned}
&(L^D \leq n\tau \leq U^D) = \\
&\quad \sum_{a^1 \in \mathcal{A}^1} (L^D \leq n\tau \leq U^D | A^1 = a^1)(A^1 = a^1) \\
&\geq \sum_{a^1 \in [L^1, U^1]} (L^D \leq A^0 + a^1 \leq U^D | A^1 = a^1)(A^1 = a^1) \\
&= \sum_{a^1 \in [L^1, U^1]} (L^D - a^1 \leq A^0 \leq U^D - a^1 | A^1 = a^1)(A^1 = a^1) \\
&= \sum_{a^1 \in [L^1, U^1]} (L^0 + (L^1 - a^1) \leq A^0 \leq U^0 + (U^1 - a^1) | A^1 = a^1)(A^1 = a^1) \\
&\geq \sum_{a^1 \in [L^1, U^1]} (L^0 \leq A^0 \leq U^0 | A^1 = a^1)(A^1 = a^1) \\
&= (L^0 \leq A^0 \leq U^0, L^1 \leq A^1 \leq U^1) \\
&\geq (L^0 \leq A^0 \leq U^0) + (L^1 \leq A^1 \leq U^1) - 1 \\
&= 1 - \alpha
\end{aligned}
$$

where the first inequality follows because $[L^1, U^1] \subseteq \mathcal{A}^1$, the second inequality is true because for all $a^1 \in [L^1, U^1]$, $L^1 - a^1 \leq 0$ and $U^1 - a^1 \geq 0$, and the third inequality follows from the Bonferroni inequality.

## APPENDIX B: TECHNICAL DETAILS FOR CHAPTER 4

*Proof of Proposition 4.1*

$$E[\hat{Y}_i^{IPW}(z;\alpha)] = E_X\left[E_{Y|X}\left\{\frac{1}{n_i}\sum_{j=1}^{n_i} E_{Z|Y,X}\left(\frac{\pi_\alpha(Z_{i(j)}|Z_{ij})1\{Z_{ij}=z\}Y_{ij}}{\Pr[Z_i=z_i|X_i]}\right)\right\}\right]$$

$$= E_X E_{Y|X}\left\{\frac{1}{n_i}\sum_{j=1}^{n_i}\sum_{\omega\in\mathcal{Z}(n_i-1)}\left(\frac{\pi_\alpha(Z_{i(j)}=\omega|Z_{ij}=z)Y_{ij}(Z_{ij}=z,Z_{i(j)}=\omega)}{\Pr[Z_{i(j)}=\omega,Z_{ij}=z]^{-1}\Pr[Z_{i(j)}=\omega,Z_{ij}=z|X_i]}\right)\right\}$$

$$= E_X\left\{E_{Y|X}\left(\frac{1}{n_i}\sum_{j=1}^{n_i}\sum_{\omega\in\mathcal{Z}(n_i-1)}\pi_\alpha(Z_{i(j)}=\omega|Z_{ij}=z)Y_{ij}(Z_{ij}=z,Z_{i(j)}=\omega)\right)\right\}$$

$$= E_X\{E_{Y|X}(\bar{Y}_i(z;\alpha))\} = \mu_{z,\alpha}$$

*Proof of Proposition 4.2*

$$E[\hat{Y}_i^{IPW}(\alpha)] = E_X E_{Y|X}\left\{\frac{1}{n_i}\sum_{j=1}^{n_i} E_{Z|Y,X}\left(\frac{\pi_\alpha(Z_i)Y_{ij}}{\Pr[Z_i=z_i|X_i]}\right)\right\}$$

$$= E_X E_{Y|X}\left\{\frac{1}{n_i}\sum_{j=1}^{n_i}\sum_{\omega\in\mathcal{Z}(n_i)}\left(\frac{\pi_\alpha(Z_i=\omega)\Pr[Z_i=\omega]Y_{ij}(Z_i=\omega)}{\Pr[Z_i=\omega|X_i]}\right)\right\}$$

$$= E_X\left\{E_{Y|X}\left(\frac{1}{n_i}\sum_{j=1}^{n_i}\sum_{\omega\in\mathcal{Z}(n_i)}\pi_\alpha(Z_i=\omega)Y_{ij}(Z_i=\omega)\right)\right\}$$

$$= E_X\{E_{Y|X}(\bar{Y}_i(\alpha))\} = \mu_\alpha$$

*Derivation of causal effect in simulation study in Section 4.4.4*

$$
\begin{aligned}
\mu_{z,\alpha} &= E[\bar{Y}_i(z;\alpha)] \\
&= E_{X,n_i}\left\{E_{Y|X,n_i}[\bar{Y}_i(z;\alpha)|X,n_i]\right\} \\
&= E_{X,n_i}\frac{1}{n_i}\sum_{j=1}^{n_i}\sum_{c=0}^{n_i-1}\binom{n_i-1}{c}\alpha^c(1-\alpha)^{n_i-1-c}\mathcal{L}^{-1}(\beta_0+\beta_1 z+\beta_2(z+c)/n_i+\beta_3 X_{ij})
\end{aligned}
$$

# BIBLIOGRAPHY


Agresti, A. and Caffo, B. (2000), "Simple and effective confidence intervals for proportions and differences of proportions result from adding two successes and two failures," *The American Statistician*, 54, 280–288.

Ali, M., Emch, M., von Seidlein, L., Yunus, M., Sack, D., Rao, M., Holmgren, J., and Clemens, J. (2005), "Herd Immunity Conferred by Killed Oral Cholera Vaccines in Bangladesh: A Reanalysis," *Lancet*, 366, 44–49.

An, W. (2010), "Bayesian propensity score estimators: incorporating uncertainties in propensity scores into causal inference," *Sociological Methodology*, 40, 151–189.

Aronow, P. (2012), "A General Method for Detecting Interference Between Units in Randomized Experiments," *Sociological Methods & Research*, 41, 3–16.

Baird, S., Garfein, R., McIntosh, C., and Özler, B. (2012), "Effect of a cash transfer programme for schooling on prevalence of HIV and herpes simplex type 2 in Malawi: a cluster randomised trial," *The Lancet*, 379, 1320–1329.

Borm, G., Melis, R., Teerenstra, S., and Peer, P. (2005), "Pseudo cluster randomization: a treatment allocation method to minimize contamination and selection bias," *Statistics in Medicine*, 24, 3535–3547.

Bowers, J., Fredrickson, M., and Panagopoulos, C. (2013), "Reasoning about Interference Between Units: A General Framework," *Political Analysis*, 21, 97–124.

Chib, S. and Hamilton, B. (2000), "Bayesian analysis of cross-section and clustered data treatment models," *Journal of Econometrics*, 97, 25–50.

Cox, D. (1958), *Planning of Experiments*, New York, NY: Wiley.

Crowley, J., Kim, Y., Lenarcic, A., Quackenbush, C., Barrick, C., Adkins, D., Shaw, G., Miller, D., de Villena, F., Sullivan, P., and W., V. (2014), "Genetics of adverse reactions to haloperidol in a mouse diallel: a drug-placebo experiment and Bayesian causal analysis," *Genetics*, 196, 321–347.

Dawid, A. (2000), "Causal inference without counterfactuals," *Journal of the American Statistical Association*, 95, 407–424.

Duflo, E. and Saez, E. (2003), "The Role of Information and Social Interactions in Retirement Plan Decisions: Evidence from a Randomized Experiment," *The Quarterly Journal of Economics*, 118, 815–842.

Fisher, R. (1935), *The Design of Experiments*, Edinburgh and London: Oliver and Boyd.

Freedman, D. and Berk, R. (2008), "Weighting regressions by propensity scores," *Evaluation Review*, 32, 392–409.

Gelman, A. (2011), "Causality and Statistical Learning," *American Journal of Sociology*, 117, 955–966.

Greenland, S. (2000), "Causal analysis in the health sciences," *Journal of the American Statistical Association*, 95, 286–289.

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

Halloran, M. and Struchiner, C. (1991), "Study Designs for Dependent Happenings," *Epidemiology*, 2, 331–338.

— (1995), "Causal Inference in Infectious Diseases," *Epidemiology*, 6, 142–151.

Hawley, W., Phillips-Howard, P., Kuile, F. t., Terlouw, D., Vulule, J., Ombok, M., Nahlen, B., Gimnig, J., Kariuki, S., Kolczak, M., et al. (2003), "Community-wide effects of permethrin-treated bed nets on child mortality and malaria morbidity in western Kenya," *American Journal of Tropical Medicine and Hygiene*, 68, 121–127.

Hii, J., Smith, T., Vounatsou, P., Alexander, N., Mai, A., Ibam, E., and Alpers, M. (2001), "Area effects of bednet use in a malaria-endemic area in Papua New Guinea," *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 95, 7–13.

Hodges, J. and Lehmann, E. (1963), "Estimates of Location Based on Rank Tests," *The Annals of Mathematical Statistics*, 34, 598–611.

Holland, P. (1986), "Statistics and Causal Inference," *Journal of the American Statistical Association*, 945–960.

Hong, G. and Raudenbush, S. (2006), "Evaluating Kindergarten Retention Policy: A Case Study of Causal Inference for Multi-level Observational Data," *Journal of the American Statistical Association*, 101, 901–910.

Hudgens, M. and Halloran, M. (2008), "Toward Causal Inference With Interference," *Journal of the American Statistical Association*, 103, 832–842.

Ichino, N. and Schündeln, M. (2012), "Deterring or displacing electoral irregularities? Spillover effects of observers in a randomized field experiment in Ghana," *The Journal of Politics*, 74, 292–307.

Kang, J. and Schafer, J. (2007), "Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data," *Statistical Science*, 22, 523–539.

LaVange, L., Durham, T., and Koch, G. (2005), "Randomization-Based Nonparametric Methods for the Analysis of Multicentre Trials," *Statistical Methods in Medical Research*, 14, 281–301.

Lehmann, E. (1998), *Nonparametrics: Statistical Methods Based on Ranks*, Upper Saddle River, NJ: Springer.

Little, R. and Rubin, D. (1989), "The analysis of social science data with missing values," *Sociological Methods & Research*, 18, 292–326.

— (2000), "Causal Effects in Clinical and Epidemiological Studies Via Potential Outcomes: Concepts and Analytical Approaches," *Annual Review of Public Health*, 21, 121–145.

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

Lunceford, J. and Davidian, M. (2004), "Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study," *Statistics in Medicine*, 23, 2937–2960.

Manski, C. (1990), "Nonparametric bounds on treatment effects," *The American Economic Review*, 80, 319–323.

Mebane Jr., W. and Sekhon, J. (2011), "Genetic optimization using derivatives: the rgenoud package for R," *Journal of Statistical Software*, 42, 1–26.

Mehta, C. and Patel, N. (2003), *StatXact 5 for Windows: Statistical Software for Exact Nonparametric Inference User Manual*, CYTEL Software Corporation.

Messina, J., Taylor, S., Meshnick, S., Linke, A., Tshefu, A., Atua, B., Mwandagalirwa, K., and Emch, M. (2011), "Population, behavioural and environmental drivers of malaria prevalence in the Democratic Republic of Congo," *Malaria Journal*, 10, 161.

Miettinen, O. and Cook, E. (1981), "Confounding: essence and detection," *American Journal of Epidemiology*, 114, 593–603.

Moulton, L., O'Brien, K., Kohberger, R., Chang, I., Reid, R., Weatherholtz, R., Hackell, J., Siber, G., and Santosham, M. (2001), "Design of a Group-Randomized Streptococcus pneumoniae Vaccine Trial," *Controlled Clinical Trials*, 22, 438–452.

Neyman, J. (1923), "On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9 (1990 Dabrowska and Speed translation)," *Statistical Science*, 5, 465–472.

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

R Core Team (2014), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria,, http://www.R-project.org/.

Rigdon, J. (2014), *RI2by2: Randomization inference for treatment effects on a binary outcome*, r package version 1.2. http://CRAN.R-project.org/package=RI2by2.

— (2015), *interferenceCI: Exact Confidence Intervals in the Presence of Interference*, r package version 1.1. http://CRAN.R-project.org/package=interferenceCI.

Robins, J. (1988), "Confidence Intervals for Causal Parameters," *Statistics in Medicine*, 7, 773–785.

Rosenbaum, P. (2001), "Effects Attributable to Treatment: Inference in Experiments and Observational Studies with a Discrete Pivot," *Biometrika*, 88, 219–231.

— (2002a), "Covariance adjustment in randomized experiments and observational studies," *Statistical Science*, 17, 286–327.

— (2002b), *Observational Studies*, New York, NY: Springer.

— (2007), "Interference Between Unites in Randomized Experiments," *Journal of the Americal Statistical Association*, 102, 191–200.

Rosenbaum, P. and Rubin, D. (1983), "The central role of the propensity score in observational studies for causal effects," *Biometrika*, 70, 41–55.

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

— (1978), "Bayesian Inference for Causal Effects: The Role of Randomization," *The Annals of Statistics*, 6, 34–58.

— (1980), "Discussion of "Randomization Analysis of Experimental Analysis of Experimental Data in the Fisher Randomization Tests," by D. Basu," *Journal of the American Statistical Association*, 75, 591–593.

— (1991), "Practical implications of modes of statistical inference for causal effects and the critical role of the assignment mechanism." *Biometrics*, 47, 1213–1234.

— (2005), "Causal Inference Using Potential Outcomes: Design, Modeling, Decisions," *The Journal of the American Statistical Association*, 100, 322–331.

Sacerdote, B. (2001), "Peer effects with random assignment: Results for Dartmouth roommates," *The Quarterly Journal of Economics*, 116, 681–704.

Santner, T. and Snell, M. (1980), "Small-Sample Confidence Intervals for p1 - p2 and p1/p2 in 2 x 2 Contingency Tables," *Journal of the American Statistical Association*, 75, 386–394.

SAS Institute Inc. (2014), *SAS Software, Version 9.3*, Cary, NC,, http://www.sas.com/.

Schwartz, S., Li, F., and Mealli, F. (2011), "A Bayesian semiparametric approach to intermediate variables in causal inference," *Journal of the American Statistical Association*, 106, 1331–1344.

Seal, K., Kral, A., Lorvick, J., McNees, A., Gee, L., and Edlin, B. (2003), "A randomized controlled trial of monetary incentives vs. outreach to enhance adherence to the hepatitis B vaccine series among injection drug users," *Drug and Alcohol Dependence*, 71, 127–131.

Sinclair, B., McConnell, M., and Green, D. (2012), "Detecting spillover effects: Design and analysis of multilevel experiments," *American Journal of Political Science*, 56, 1055–1069.

Sobel, M. (2006), "What Do Randomized Studies of Housing Mobility Demonstrate?: Causal Inference in the Face of Interference," *Journal of the Americal Statistical Association*, 101, 1398–1407.

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

Stewart, W. (2002), "Groundhog day: cause and effect and the primary importance of the finite population induced by randomization," *Journal of Biopharmaceutical Statistics*, 12, 93–105.

Tchetgen Tchetgen, E. and VanderWeele, T. (2012), "On Causal Inference in the Presence of Interference," *Statistical Methods in Medical Research*, 21, 55–75.

Thulin, M. (2014), "Coverage-adjusted Confidence Intervals for a Binomial Proportion," *Scandinavian Journal of Statistics*, 41, 291–300.

Toulis, P. and Kao, E. (2013), "Estimation of Causal Peer Influence Effects," in *Proceedings of The 30th International Conference on Machine Learning*, pp. 1489–1497.