CAUSAL INFERENCE WITH PARTIAL INTERFERENCE

Jenna Landy
jlandy@g.harvard.edu

Christian Testa
ctesta@hsph.harvard.edu

Daniel Xu
danielxu@g.harvard.edu

May 14, 2024

ABSTRACT

Standard causal inference methods often make use of the assumption of no interference, which states that an observation's treatment does not affect any other observation's outcome. Though often reasonable, this assumption breaks down in many settings, including those involving infectious diseases and social networks. In this review, we provide an overview of popular methods developed to handle partial interference, where we assume that individuals may only interfere with other individuals in their defined group. After defining causal estimands relevant to the interference setting, we present estimators for these estimands, as well as methods for inference and constructing confidence intervals that have been developed in recent years. Through simulations under partial interference, we demonstrate the need for these estimators over traditional treatment effect estimators that otherwise ignore interference. Lastly, we conclude by discussing some limitations of existing methods and areas of future research.

1 Introduction

Standard causal inference methods use the stable unit treatment value (SUTVA) assumption, which includes the assumption that one observation's treatment does not affect any other observation's outcome. This is known as the assumption of *no interference* [Cox, 1958, Rubin, 1980]. While this assumption may be reasonable in many settings, there are clear violations to the no interference assumption, for example in the context of infectious diseases, social networks, or interventions that can only be group-randomized [Hudgens and Halloran, 2008]. Classic examples of causal inference under interference include estimating the effect of vaccination on infectious disease outcomes when vaccination among that individuals' contacts may also induce a protective effect, violating SUTVA. Other examples from econometrics include the effects of housing mobility programs and their spillover effects Sobel [2006], or in school-based interventions when students who receive an intervention may help or share information with students who do not receive the intervention.

The concept of interference between sample units dates back to 1616 when Sir Ronald Ross defined "dependent happenings", where an event's rate of occurrence depends on the number of individuals already affected [Ross, 1916]. Definitions of direct versus indirect effects began popping up in the early '90s [Halloran et al., 1991, Struchiner et al., 1990, Halloran and Struchiner, 1995]. However, the discussion of these topics within the scope of causal inference is much more recent, with main works appearing in the last 20 years [Tchetgen and VanderWeele, 2012, Hudgens and Halloran, 2008, Rosenbaum, 2007].

Depending on the application, interference may be treated as a nuisance to be avoided when possible [Rosenbaum, 2007] or as an important, sometimes biological, concept that we wish to understand by estimating direct and indirect effects. Within the second group, Halloran and Hudgens [2016] broadly divides up causal methods for interference into the following categories: (1) partial interference, where distinct blocks or groups are identified such that interference can occur within, but not across, groups, (2) general or network interference, where interference is less restricted with a graph or network structure, and (3) studies of infectiousness and contagion. **In this paper, we focus on (1) partial interference in the context of vaccine trials, where a person's level of infection depends on who else in their block or group is vaccinated (i.e., herd immunity)**. We refer readers to Halloran and Hudgens [2016] for overviews of and references for the other two branches of study.

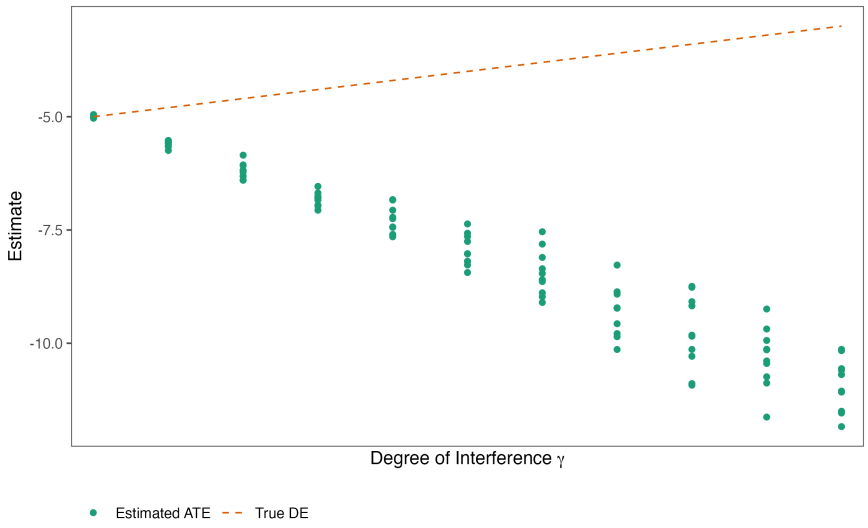

Figure 1: Naive estimates of average treatment effect (ATE) compared to the true population direct effect (DE) as degree of interference increases.

The rest of this paper is structured as follows. In Section 2, we introduce notation for group-randomized trials and define potential outcomes in this framework. In Section 3, we define total, overall, direct, and indirect effects of interest and interpret them in the context of a vaccine trial. In Section 4, we recap standard and introduce new identification assumptions and identify the causal effects of interest. Section 5 defines commonly used estimators for the causal effects of interest as well as estimators of their variance. Section 6 uses a hypothetical data generating function and simulated data to express the difference between two definitions of the direct effect. Finally sections 7, 8, and 9 discuss applications where these methods have been used in practice, limitations of these methods, and recent works and current areas of research.

1.1 Hypothetical Vaccine Trial Example

Throughout this paper, we describe quantities in the context of a theoretical vaccine trial adapted from Hudgens and Halloran [2008]. Treatment is a vaccination compared to a no-treatment placebo. The outcome is a continuous measure of the severity of infection within the next year. The groups are defined as five geographically separate regions, each of 100 observations. Three groups are assigned to a randomization strategy that randomizes treatment to 30 individuals, and the other two groups are assigned to a randomization strategy that randomizes treatment to 50 individuals.

1.2 Motivating Example

We start with a motivating example of the importance of the interference assumption and how standard methods fail when the assumption is broken but ignored. Suppose that in the context of the hypothetical vaccine trial, we wish to estimate the effect of an individual's vaccination status on the severity of their disease. If we ignore any possibility of interference and incorrectly say that the no interference assumption is met, we would compute the ATE and interpret it as a direct effect.

Across varying degrees of interference, we simulated 50 datasets according to the setup in Section 6.1 and use the difference in means approach to estimate the ATE. Figure 1 shows that as the degree of interference γ increases, these naive estimates fall far away from the true direct effect.

One can imagine a more extreme scenario where ignoring interference may mean estimating a null effect or estimating an effect of the wrong direction. On the other hand, as in the example above, the magnitude of the effect may be far overestimated, which would be misleading, for example, in clinical trial results.

For another example, it may be of significant importance to regulatory agencies like the Food and Drug Administration whether the effect of a drug is largely through direct effects on individuals or through a composite of direct and spillover/interference effects. If there are significant spillover effects, and due to variation in contact patterns and

social network structures, we may expect large effect heterogeneity due to effect modification according to the level of interference.

Clearly, it is necessary to correctly handle interference when it is present by estimating direct and indirect effects separately.

2 Definitions

In this paper, we adapt the notation of Hudgens and Halloran [2008] with a few changes to match the notation used in class.

2.1 Partial Interference

We suppose there are N blocks of units or groups of individuals such that interference may occur within each block. That is, the treatment of one individual may affect the outcome of themselves along with the outcomes of other observations in their block, but may not affect the outcomes of individuals outside of their block. This is called the partial interference assumption [Sobel, 2006], the violation of which is called contamination [Hudgens and Halloran, 2008]. Block i holds n_i individuals, and $\mathbf{A}_i = (A_{i1}, A_{i2}, ..., A_{in_i})$ is the vector of treatment assignments, or the treatment program, within block i. $\mathbf{A}_{i(j)}$ is the subvector of treatments in block i excluding observation j. The concept of partial interference is demonstrated in Figure 2.

As in Hudgens and Halloran [2008], we assume each block may have its own treatment assignment method, which we denote as S_i. We visualize this as a two-step randomization procedure where each group is assigned to a treatment method then each individual is assigned treatment according to their group's method. We define two possible treatment assignment mechanisms, ψ and ϕ such that $S_i \in \{\psi, \phi\}$.

In the context of our theoretical vaccine trial, the groups are the geographically separated regions. Under the partial interference assumption, individuals do not come into contact with individuals outside of their group. Examples of potential contamination include people traveling for work or visiting distant relatives. The different treatment mechanisms are ψ, which randomizes treatment to 30 of 100 individuals, and ϕ, which randomizes treatment to 50 of 100 individuals.

2.2 Group (Two-Stage) Randomization versus Cluster Randomization

When conducting a study under the partial interference assumption, one must decide between group (also called two-stage) randomization, which we focus on in this paper, and cluster randomization. The difference lies in the treatment assignment strategy, and the choice makes implications on the ability to estimate direct and indirect effects.

Cluster randomized trials assign treatment at the cluster level such that every observation in a cluster is treated or every observation is untreated. This is often due to limitations, such as a policy that can only be implemented at the village level or a teaching strategy that can only be implemented at the classroom level [Dron et al., 2021]. In cluster randomized trials, only the overall effect on each cluster can be identified and estimated, not the direct and indirect effects within clusters [Benitez et al., 2023].

Group randomized trials assign a cluster to a particular randomization strategy. Individuals are then assigned to treatment or no treatment according to its group's randomization strategy. This is also called two-stage randomization [Halloran and Hudgens, 2016]. In a group randomized trial, both direct and indirect effects can be estimated as long as all randomization strategies are *mixed assignment strategies*, meaning they assign at least one individual to each of the treated and untreated groups [Hudgens and Halloran, 2008].

In the context of vaccinations, it is important to understand separately the direct and indirect effects. These effects can point to how protective *my own vaccination status* is versus *my community's vaccination status*. For this reason, we focus on group randomized trials.

2.3 Finite Sample versus Bernoulli Group-Randomization

Within group-randomized trials, there are two types of randomization strategies: finite sample (also called type A) and Bernoulli randomization (also called type B). In the finite sample approach, a fixed number of individuals are treated. This causes a dependence in treatment assignments: if individual 1 is assigned treatment, that decreases the probability that individual 2 is assigned treatment. In the Bernoulli approach, individuals are assigned treatment independently

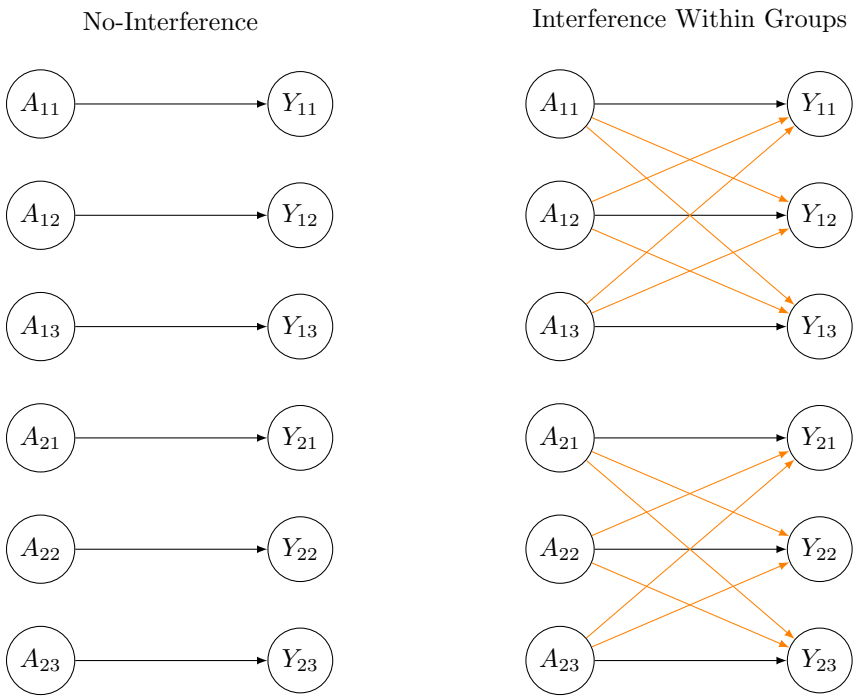

Figure 2: Conceptual diagram illustrating the difference between conventional no-interference treatment and outcomes at the individual level and with interference within groups. A_{ij} represents the exposure/treatment status for the jth individual who belongs to the ith group, and similarly for the outcomes Y_{ij}. Here there are 3 groups with 3 individuals in each. The assumption of partial interference is encoded by the fact that on the right-hand panel there are no orange arrows going across groups.

according to a Bernoulli distribution with a fixed probability of treatment. These parameterizations are formalized in Tchetgen and VanderWeele [2012], again using ϕ and ψ as possible treatment assignment strategies.

For a finite-sample (type A) randomization strategy, the strategy ϕ defines the number of observations who should be assigned treatment K_{ϕ,n_i} given a sample size n_i. For example ϕ might be the assignment strategy that assigns half of samples to be treated. All possible treatment programs $\mathbf{A}_i$ conditional on $\sum_{j=1}^{n_i} A_{ij} = K_{\phi,n_i}$ are given equal probability with probability mass function

$$P_i(\mathbf{A}_i|\phi) = I\left(\sum_{j=1}^{n_i} A_{ij} = K_{\phi,n_i}\right) \frac{1}{\binom{n_i}{K_{\phi,n_i}}}$$

where $I(\cdot)$ is the indicator function.

For a Bernoulli (type B) randomization strategy, the strategy ϕ defines the probability of each individual being treated p_ϕ. For example, ϕ might be the assignment strategy that assigns treatment with probability 0.5. Here, treatment assignments are independent, so treatment programs $\mathbf{A}_i$ are assigned according to the probability mass function

$$P_i(\mathbf{A}_i|\phi) = \prod_{j=1}^{n_i} (p_\phi)^{A_{ij}} (1 - p_\phi n_i)^{1-A_{ij}}$$

for a given sample size n_i.

We use the notation ϕ^n or ψ^n to define the set of all possible treatment assignment programs $\mathbf{A}_i$ to n individuals from assignment strategy ϕ or ψ, respectively.

2.4 Potential Outcome Framework with Interference

The standard counterfactual notation is extended in the presence of interference to depend on the entire vector $\mathbf{a}_i$ rather than a single observation's a_{ij}. In this setting $Y_{ij}(\mathbf{a}_i) = Y_{ij}(A_{i1} = a_{i1}, ..., A_{ij} = a_{ij}, A_{in_i} = a_{in_i})$ is the potential outcome for observation j of block i when the treatment program for block i is set to $\mathbf{a}_i$. We often write this in a form that isolates the treatment of the individual of interest: $Y_{ij}(\mathbf{a}_i) = Y_{ij}(\mathbf{a}_{i(j)}, A_{ij} = a_{ij})$, where $\mathbf{a}_{i(j)}$ is the vector $\mathbf{a}_i$ excluding entry j. This new counterfactual notation makes it clear that the outcome of observation j in block i can change with (1) its own treatment, a_{ij}, and (2) the treatments of other individuals in block i, $\mathbf{a}_{i(j)}$.

The individual average potential outcome of observation j in block i when treatment $A_{ij} = a$ is defined as the average over all counterfactuals where $A_{ij} = a$, that is, the average over all possible values of the other treatment assignments in block i:

$$\begin{aligned}\bar{Y}_{ij}(a|\psi) &= E_\psi[Y_{ij}(\mathbf{a}_i)|A_{ij} = a] \\ &= \sum_{\omega \in \psi^{n_i - 1}} Y_{ij}(\mathbf{A}_{i(j)} = \omega, A_{ij} = a) \cdot Pr_\psi(\mathbf{A}_{i(j)} = \omega | A_{ij} = a)\end{aligned}$$

The randomization strategy makes an important difference on the average potential outcome. In finite sample (type A) randomization, the treatments $A_{i(j)}$ and A_{ij} depend on each other. In Bernoulli (type B) randomization, they are independent such that $Pr_\psi(\mathbf{A}_{i(j)} = \omega | A_{ij} = a) = Pr_\psi(\mathbf{A}_{i(j)} = \omega)$.

The average potential outcome under treatment a in block i is the individual average potential outcome averaged across individuals. Similarly, the population average potential outcome is the average over blocks.

$$\begin{aligned}\bar{Y}_i(a|\psi) &= \frac{1}{n_i} \sum_{j=1}^{n_i} \bar{Y}_{ij}(a|\psi) \\ \bar{Y}(a|\psi) &= \frac{\sum_{i=1}^{N} \bar{Y}_i(a|\psi)}{\sum_{i=1}^{N} I(S_i = \psi) I(S_i = \psi)}\end{aligned}$$

Contrast these definitions with the marginal individual average potential outcome, defined solely as a function of the group assignment (and not the individual's treatment assignment as in above):

$$\begin{aligned}\bar{Y}_{ij}(\psi) &= \mathbb{E}_\psi[Y_{ij}(\mathbf{a}_i)] \\ &= \sum_{\omega \in \psi^{n_i}} Y_{ij}(\omega) \cdot Pr_\psi(\mathbf{A}_i = \omega).\end{aligned}$$

This is a marginal quantity in the sense that it is average potential outcome for an individual marginalized over all possible treatment assignments for a group assigned to ψ. Similar to before, we can also define marginal group and population average potential outcomes as follows:

$$\begin{aligned}\bar{Y}_i(\psi) &= \frac{1}{n_i} \sum_{j=1}^{n_i} \bar{Y}_{ij}(a|\psi) \\ \bar{Y}(\psi) &= \frac{1}{N} \sum_{i=1}^{N} \bar{Y}_i(\psi).\end{aligned}$$

In the context of our hypothetical vaccine trial, $\bar{Y}_{ij}(a|\psi)$ is the expected severity of infection within one year for individual j within group i under treatment assignment a assuming the block has been randomized according to mechanism ψ, that is, assuming treatment was randomized to 30 of 100 individuals in group i. Likewise, $\bar{Y}_{ij}(\psi)$ can be interpreted as the expected severity of infection within one year for individual j within group i that has been randomized according to mechanism ψ.

3 Causal Estimands

There are four types of causal estimands of interest in the context of partial interference: direct effects, indirect or spillover effects, total effects, and overall effects. These are represented visually in Figure 3 and mathematically defined in the following sub-sections.

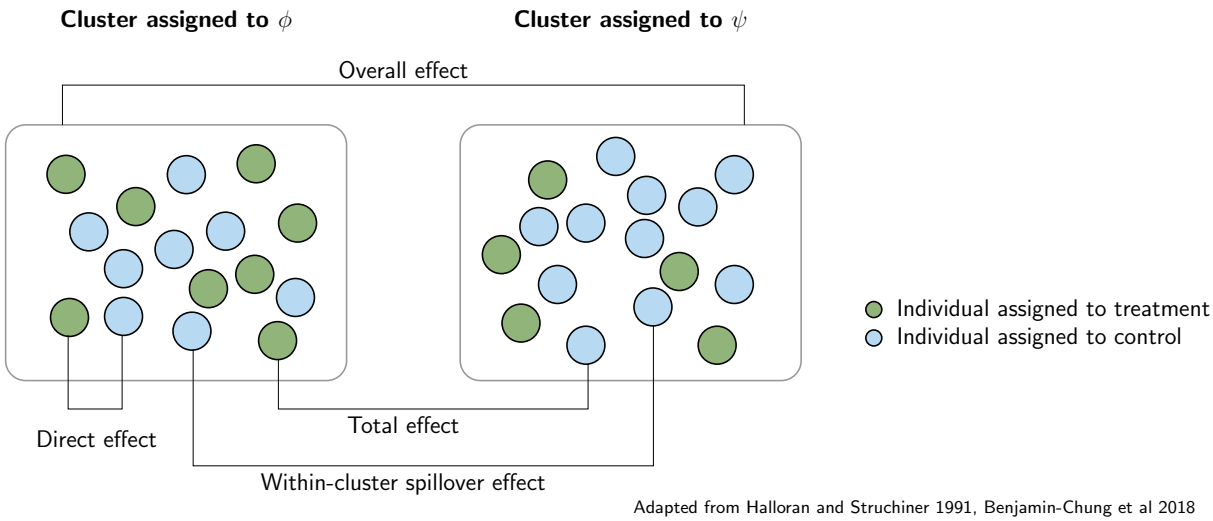

Figure 3: Adapted from Halloran et al. [1991], Benjamin-Chung et al. [2018], this figure shows the concepts behind each of the causal estimands we examine: the direct effect should reflect the effects of different treatment assignments within a group or randomization strategy (within a group, any spillover effect from others are present in equal amounts for both the treatment and controlled unit). The total effect should capture both the impact of a treatment on an individual and the spillover effects from the treatment strategy/regime. The within-cluster spillover effect should represent the effects of only spillover, while the individual treatment assignment is held fixed. Finally, overall effects should capture the group-level differences in outcomes resulting from different treatment strategies assigned at the group-level.

3.1 Direct (Protective) Effects

A direct effect is defined as the effect of a treatment of an individual on their own outcome holding all else constant, that is, holding other individuals' treatments at particular values [Halloran et al., 1991, Halloran and Struchiner, 1995]. These are often called "protective effects" in the context of vaccinations where we expect the treatment to be protective against, or reduce the risk or severity of, disease. The protective effect, then, is the benefit a treatment has directly on the person receiving it [Halloran and Hudgens, 2016].

The individual direct causal effect measures the difference in counterfactuals for observation j in block i while holding treatment assignments constant for all other observations in block i. Averaging over all possible treatment assignments for the other observations gives the individual average direct effect. Further averaging this over individuals in each block gives the group average direct effect. Finally, averaging over these gives the population average direct effect.

- Individual direct effect $DE_{ij}(\mathbf{a}_{i(j)}) = Y_{ij}(\mathbf{a}_{i(j)}, A_{ij} = 1) - Y_{ij}(\mathbf{a}_{i(j)}, A_{ij} = 0)$
- Individual average direct effect $\overline{DE}_{ij}(\psi) = \bar{Y}_{ij}(1|\psi) - \bar{Y}_{ij}(0|\psi)$
- Group average direct effect $\overline{DE}_i(\psi) = \bar{Y}_i(1|\psi) - \bar{Y}_i(0|\psi) = \frac{1}{n_i}\sum_{j=1}^{N} \overline{DE}_{ij}(\psi)$
- Population average direct effect $\overline{DE}(\psi) = \bar{Y}(1|\psi) - \bar{Y}(0|\psi) = \frac{1}{N}\overline{DE}_i(\psi)$

If there is no interference, then the group average direct effect is equivalent to the typical average treatment effect (ATE).

In the context of our hypothetical vaccine trial, the population average direct effect $\overline{DE}(\psi)$ is the expected difference in level of infection when an individual is vaccinated versus unvaccinated, assuming the population is assigned to randomization strategy ψ such that 30 of 100 individuals are vaccinated.

3.2 Indirect (Spillover) Effects

An indirect effect is defined as the effect of other individuals' treatment assignments on the outcome of an untreated individual. The definition here is with respect to two arbitrary intervention programs on the other individuals, $\mathbf{a}_{i(j)}$ and $\mathbf{a}'_{i(j)}$. Sometimes, the contrast between $\mathbf{a}_{i(j)}$ being all treated versus $\mathbf{a}'_{i(j)}$ being all untreated is of particular interest

[Halloran and Struchiner, 1995]. These are often called 'spillover effects" in the context of vaccine trials, where the indirect effect is the spillover benefit that a treatment has on those who did not directly receive it.

Individual indirect effects compare the counterfactuals for observation j in block i when its own treatment is held at zero, but the treatments of other individuals in its block vary between two treatment assignments or treatment assignment mechanisms. Again, by sequentially averaging effects we can get the individual average indirect effect, the group average indirect effect, and the population average indirect effect.

- Individual indirect effect $IE_{ij}(\mathbf{a}_{i(j)}, \mathbf{a'}_{i(j)}) = Y_{ij}(\mathbf{a}_{i(j)}, A_{ij} = 0) - Y_{ij}(\mathbf{a}'_{i(j)}, A_{ij} = 0)$
- Individual average indirect effect $\overline{IE}_{ij}(\phi, \psi) = \bar{Y}_{ij}(0|\phi) - \bar{Y}_{ij}(0|\psi)$
- Group average indirect effect $\overline{IE}_i(\phi, \psi) = \bar{Y}_i(0|\phi) - \bar{Y}_i(0|\psi) = \frac{1}{n_i} \sum_{j=1}^{n_i} \overline{IE}_{ij}(\phi, \psi)$
- Population average indirect effect $\overline{IE}(\phi, \psi) = \bar{Y}(0|\phi) - \bar{Y}(0|\psi) = \frac{1}{N} \sum_{i=1}^{N} \overline{IE}_i(\phi, \psi)$

If there is no interference, then the individual indirect effect is 0 for all individuals, and thus the group and population average indirect effects are also 0.

In the context of our hypothetical vaccine trial, the populate average indirect effect $\overline{IE}(\phi, \psi)$ is the expected difference in level of infection of an unvaccinated individual comparing the world where treatment mechanism ϕ was used, where 50 of 100 individuals are vaccinated, to the world where treatment mechanism ψ was used, where 30 of 100 individuals are vaccinated.

3.3 Total Effects

The individual total effect measures the difference in the potential outcomes of observation j in block i due to (1) changes in its own treatment and (2) changes in the treatments of other individuals in block i. In particular, this contrast is comparing $A_{ij} = 0$ and a given specification of other treatments $\mathbf{a}_{i(j)} \sim \phi$ versus $A_{ij} = 1$ and a different specification of other treatments $\mathbf{a}'_{i(j)} \sim \psi$.

- Individual total effect $TE_{ij}(\mathbf{a}, \mathbf{a}') = Y_{ij}(\mathbf{a}_{i(j)}, A_{ij} = 1) - Y_{ij}(\mathbf{a}'_{i(j)}, A_{ij} = 0)$
- Individual average total effect $\overline{TE}_{ij}(\phi, \psi) = \bar{Y}_{ij}(1|\phi) - \bar{Y}_{ij}(0, \psi)$
- Group average total effect $\overline{TE}_i(\phi, \psi) = \bar{Y}_i(1|\phi) - \bar{Y}_i(0, \psi) = \frac{1}{n_i} \sum_{j=1}^{n_i} \overline{TE}_{ij}(\phi, \psi)$
- Population average total effect $\overline{TE}(\phi, \psi) = \bar{Y}(1|\phi) - \bar{Y}(0|\psi) = \frac{1}{N} \sum_{i=1}^{N} \overline{TE}_i(\phi, \psi)$

Total effects can be decomposed into direct and indirect effects: $TE_{ij}(\mathbf{a}, \mathbf{a}') = DE_{ij}(\mathbf{a}') + IE_{ij}(\mathbf{a}, \mathbf{a}')$ and $\overline{TE}(\phi, \psi) = \overline{DE}(\psi) + \overline{IE}(\phi, \psi)$. If there is no interference, the total effect is equivalent to the direct effect because indirect effects are 0.

In the context of our hypothetical vaccine trial, the population average total effect $\overline{TE}(\phi, \psi)$ is the difference in severity of infection in an individual comparing the worlds where (1) they are treated and others in their group are assigned treatment such that 50 of 100 in their block are treated versus (2) they are untreated and others in their group are assigned treatment such that 30 of 100 in their block are treated.

3.4 Overall Effects

An overall individual effect is defined as the difference in the potential outcomes of observation j in block i due to the change from one treatment assignment vector $\mathbf{a}_i$ and a second treatment assignment vector $\mathbf{a}'_i$. Note here that $OE_{ij}(\mathbf{a}_i, \mathbf{a}'_i)$ will equal $TE_{ij}(\mathbf{a}_i, \mathbf{a}'_i)$ if $A_{ij} = 1$ in $\mathbf{a}_i$ and $A_{ij} = 0$ in $\mathbf{a}'_i$. However, the connection between overall effects and total effects does not continue beyond the individual effects.

The individual average overall effect is difference in the potential outcomes of observation j in block i between the world under assignment mechanism ϕ versus that under assignment mechanism ψ. Averaging over all observations in group i gives the group average overall causal effect, and averaging these over groups gives the population average overall effect.

- Individual overall effect $OE_{ij}(\mathbf{a}_i, \mathbf{a}'_i) = Y_{ij}(\mathbf{a}_i) - Y_{ij}(\mathbf{a}'_i)$
- Individual average overall effect $\overline{OE}_{ij}(\phi, \psi) = \bar{Y}_{ij}(\phi) - \bar{Y}_{ij}(\psi)$

- Group average overall effect $\overline{OE}_{ij}(\phi, \psi)$
- Population average overall effect $\overline{OE}(\phi, \psi) = \bar{Y}(\phi) - \bar{Y}(\psi)$

In the context of our hypothetical vaccine trial, the population average overall $\overline{OE}(\phi, \psi)$ is the difference in level of infection in an individual comparing the worlds where (1) 50 of 100 individuals in their group are assigned treatment (2) 30 of 100 individuals in their group are assigned treatment.

3.5 Redefining the Direct Effect

VanderWeele and Tchetgen Tchetgen [2011] identifies an important limitation of the so-called "direct effect" proposed in Hudgens and Halloran [2008] and presented in Section 3.1. Under the finite-sample setting, assignment strategy ψ specifies that exactly K_{ϕ,n_i} individuals are treated. This means that the setting where individual ij is treated under randomization strategy ψ has one fewer other treated individuals than the setting where individual ij is *untreated* under randomization strategy ψ. Thus, the so-called "direct effect" includes the effect of an individual being vaccinated *and* of one fewer other individuals in its group being vaccinated. To address this, VanderWeele and Tchetgen Tchetgen [2011] define the following direct, indirect, and overall effects. We use the superscript "*" to differentiate these estimands from the standard ones.

We start by defining the new individual average potential outcome. The first definition clarifies the key difference between this and the standard definition: regardless of the treatment value a of the individual in question, the other individuals are held at treatment as if the individual in question was given treatment a'. This removes the effect of "one fewer individuals treated" from the direct effect. The second definition marginalizes over possible values of a'.

$$\bar{Y}^*_{ij}(a|\psi, a') = \sum_{\omega \in \psi^{n_i - 1}} Y_{ij}(\mathbf{a}_{i(j)} = \omega, a_{ij} = a) Pr_\psi(\mathbf{A}_i = \omega | a_{ij} = a')$$

$$\bar{Y}^*_{ij}(a|\psi) = \sum_{\omega \in \psi^{n_i}} Y_{ij}(\mathbf{a}_{i(j)} = \omega_{(j)}, a_{ij} = a) Pr_\psi(\mathbf{A}_i = \omega)$$

$$\bar{Y}^*_{ij}(\psi) = \sum_{\omega \in \psi^{n_i}} Y_{ij}(\mathbf{a}_i = \omega) Pr_\psi(\mathbf{A}_i = \mathbf{s})$$

The individual average direct, indirect, and overall effects are defined below in terms of the new average counterfactuals. Note that the group and population average versions of these estimands can be built up in the same way as before.

- Individual average direct effect $\overline{DE}^*_{ij}(\psi, a) = \bar{Y}^*_{ij}(1|\psi, a) - \bar{Y}^*_{ij}(0|\psi, a)$
- Individual average indirect effect $\overline{IE}^*_{ij}(\phi, \psi) = \bar{Y}^*_{ij}(0|\phi) - \bar{Y}^*_{ij}(0|\psi)$
- Individual average overall effect $\overline{OE}^*_{ij}(\phi, \psi) = \bar{Y}^*_{ij}(\phi) - \bar{Y}^*_{ij}(\phi)$

Another benefit of this formulation is that the decomposition with the newly defined direct effect is no longer a decomposition of the total effect but of the overall effect, which the authors claim is the effect most useful in policy decision making. This new definition of a direct effect is only identifiable with Bernoulli (type B) randomization.

However convincing this paper was, the standard definition still prevails. VanderWeele and Tchetgen Tchetgen [2011] has only received 75 citations since 2011 in contrast to the 884 citations of the paper with the original direct effect definition, Hudgens and Halloran [2008], and the 508 citations of Tchetgen and VanderWeele [2012], which continues to use the original definition. In Section 6, we compare the two versions of direct effects in the context of our hypothetical vaccine trial.

4 Assumptions

The assumptions necessary to identify causal parameters in the presence of interference in randomized settings are not entirely surprising: they represent restrictions on how strong the interference can be (as in, interference is allowable within study-groups but not across study-groups; Assumption 1: partial interference), how subtle the interference can mechanism be (as in, the effect of interference should be entirely captured simply by the proportion of treated within each study-group; Assumption 2: stratified interference), and versions of the standard positivity and conditional exchangeability/randomization assumptions extended to the group-randomized setting (Assumptions 3 and 4). As this work builds on the potential outcomes framework by relaxing the "SUTVA" assumptions, the work also assumes a form

of consistency: that $Y_{ij} = Y_{ij}(a|S_i)$ whenever $A_{ij} = a$ and recalling that S_i is the treatment assignment strategy ψ or ϕ.

Assumption 1 (Partial interference), Tchetgen and VanderWeele [2012], Hudgens and Halloran [2008]: Let $\mathbf{S} \equiv (S_1, ..., S_n)$ denote the first stage of randomization group assignments with $S_i \in \{\psi, \phi\}$. Let η denote the parameterization for the distribution of S and let $C = \sum_i I(S_i = \psi)$ denote the number of groups assigned ψ. Then $\{\eta, \phi, \psi\}$ are assumed to be finite sample (type A) parameterized.

Assumption 2 (Stratified interference), Hudgens and Halloran [2008]: For any treatment assignment ψ, $Y_{ij}(\mathbf{A}_{i(j)} = \mathbf{a}_{i(j)}, A_{ij} = a'_{ij}) = Y_{ij}(\mathbf{A}_{i(j)} = \mathbf{a}'_{i(j)}, A_{ij} = a'_{ij})$ for all $\mathbf{a}_{i(j)}, \mathbf{a}'_{i(j)} \in \psi^{n_i - 1}$. Here, $\psi^{n_i - 1}$ represents the set of possible treatment programs such that there are $K_{\psi, n_i - 1}$ individuals who received the intervention for group i. This means that the indirect effect doesn't depend on *who* else in a group is assigned treatment, just *how many* are assigned treatment.

Assumption 3 (Group-Level Randomization), Tchetgen and VanderWeele [2012]: Under randomization, group counterfactuals are independent of treatment such that $\mathbf{A}_i \perp \mathbf{Y}_i(\mathbf{a})$ for all $\mathbf{a}$. This extends to group-level conditional randomization in an observation study with $\mathbf{A}_i \perp \mathbf{Y}_i(\mathbf{a})|\mathbf{L}_i$ for all $\mathbf{a}$.

Assumption 4 (Group-Level Positivity), Tchetgen and VanderWeele [2012]: For $i = 1, ..., N$ we assume that conditional on $\mathbf{L}_i$, we have that for all $\mathbf{a}_i \in \psi^{n_i}$, $P(\mathbf{A}_i = \mathbf{a}_i \mid \mathbf{L}_i) > 0$, and the same for ϕ. In settings where covariates are not of interest, we assume that for all $\mathbf{a}_i \in \psi^{n_i}$, $P(\mathbf{A}_i = \mathbf{a}_i) > 0$, and the same for ϕ.

5 Identification, Estimation, and Inference

5.1 Identification

We assume a randomized trial with no confounding. The following identification implicitly uses assumption 1 by assuming our partial interference definitions are correct, and it explicitly uses assumptions 3 and 4.

$$\begin{aligned}
\bar{Y}_{ij}(a|\psi) &= E_\psi[Y_{ij}(\mathbf{a}_i)|A_{ij} = a] \text{ by definition and positivity of } A_{ij} \\
&= E_\psi[Y_{ij}(\mathbf{a}_i)|\mathbf{A}_{i(j)} = \mathbf{a}_{i(j)}, A_{ij} = a] \text{ randomization and positivity on } \mathbf{A}_{i(j)} \\
&= E_\psi[Y_{ij}|\mathbf{A}_{i(j)} = \mathbf{a}_{i(j)}, A_{ij} = a] \text{ consistency}
\end{aligned}$$

The identification of all estimands in Section 3 follow directly from this.

5.2 Estimation

5.2.1 Plug-In Estimators for Finite Sample Randomization

Plug in estimators for each type of effect are defined assuming ϕ and ψ are finite sample randomization strategies.

Let

$$\hat{Y}_i(a|\psi) = \frac{\sum_{j=1}^{n_i} I(A_{ij} = a) Y_{ij}}{\sum_{j=1}^{n_i} I(A_{ij} = a)} = \frac{1}{K_{\psi, n_i}} \sum_{j=1}^{n_i} Y_{ij} I(A_{ij} = a), \text{ assuming } S_i = \psi,$$

$$\hat{Y}(a|\psi) = \frac{\sum_{i=1}^{N} \hat{Y}_i(a|\psi) I(S_i = 1)}{\sum_{i=1}^{N} I(S_i = 1)}$$

$$\hat{Y}_i(\psi) = \frac{\sum_{j=1}^{n_i} Y_{ij}}{n_i}, \text{ assuming } S_i = \psi, \text{ and}$$

$$\hat{Y}(\psi) = \frac{\sum_{i=1}^{n_i} \hat{Y}_i(\psi) I(S_i = \psi)}{\sum_{i=1}^{N} I(S_i = \psi)}.$$

and analogously for $\hat{Y}_i(a|\phi)$, $\hat{Y}(a|\phi)$ and $\hat{Y}(\phi)$. Then Hudgens and Halloran [2008] propose the following estimators for the group average direct effect, population direct effect, population indirect effect, population total effect, and population overall effect:

$$\widehat{DE}_i(\psi) = \hat{Y}_i(0|\psi) - \hat{Y}_i(1|\psi),$$

$$\widehat{DE}(\psi) = \hat{Y}(0|\psi) - \hat{Y}(1|\psi),$$

$$\widehat{IE}(\phi,\psi) = \hat{Y}(0|\phi) - \hat{Y}(0|\psi),$$
$$\widehat{TE}(\phi,\psi) = \hat{Y}(1|\phi) - \hat{Y}(1|\psi),$$
$$\widehat{OE}(\phi,\psi) = \hat{Y}(\phi) - \hat{Y}(\psi),$$

which they show to be unbiased under *Assumption 1*.

Proof of Unbiasedness Result for $\widehat{DE}$, $\widehat{IE}$, $\widehat{TE}$ (based on Hudgens and Halloran [2008]): Notice that the estimators $\widehat{DE}$, $\widehat{IE}$, and $\widehat{OW}$ are constructed out of the estimators $\hat{Y}(a|\psi)$ and $\hat{Y}(a|\phi)$. Without loss of generality, we will show the unbiasedness of $\hat{Y}(1|\psi)$ with $a = 1$. The same proof follows for each of the $\hat{Y}(0|\psi)$, $\hat{Y}(1|\phi)$, and $\hat{Y}(0|\phi)$. Under the assumption of partial interference and finite sample randomization strategy ψ, K_{ψ,n_i} is the number of individuals assigned treatment in a group with sample size n_i. We start with the definition of

$$\mathbb{E}\left(\hat{Y}_i(1|\psi) \mid S_i = \psi\right) = \sum_{\omega\in\psi^{n_i}} \left(\frac{1}{K_{\psi,n_i}} \sum_{j=1}^{n_i} Y_{ij}(\mathbf{a}) I(a_j = 1)\right) Pr_\psi(\mathbf{A}_i = \omega),$$

where ψ^{n_i} is the set of possible assignments of $\mathbf{A}_i$ such that $\sum_j^{n_i} A_{ij} = K_{\psi,n_i}$. Since the terms where $a_j = 0$ do not contribute to the sum, we can equivalently write

$$\begin{aligned}\mathbb{E}\left(\hat{Y}_i(1|\psi)|S_i = \psi\right) &= \frac{1}{K_{\psi,n_i}} \sum_{j=1}^{n_i} \sum_{\omega\in\psi^{n_i-1}} Pr_\psi(\mathbf{A}_{i(j)} = \omega, A_{ij} = 1) \cdot Y_{ij}(A_{ij} = 1, \mathbf{A}_{i(j)} = \omega) \\ &= \frac{1}{K_{\psi,n_i}} \sum_{j=1}^{n_i} \sum_{\omega\in\psi^{n_i-1}} Pr_\psi(\mathbf{A}_{i(j)} = \omega|A_{ij} = 1) \cdot Pr(A_{ij} = 1) \cdot Y_{ij}(A_{ij} = 1, \mathbf{A}_{i(j)} = \omega).\end{aligned}$$

Under the assumption of finite sample (type A) group randomization, we have that $Pr_\psi(A_{ij} = 1) = K_{\psi,n_i}/n_i$, which implies that

$$\begin{aligned}\mathbb{E}\left(\hat{Y}_{ij}(1|\psi)|S_i = \psi\right) &= \frac{1}{n_i} \sum_{j=1}^{n_i} \sum_{\omega\in\psi^{n_i-1}} Pr_\psi(\mathbf{A}_{i(j)} = \omega|\mathbf{A}_{ij} = 1) \cdot Y_{ij}(A_{ij} = 1, \mathbf{A}_{i(j)} = \omega) \\ &= \bar{Y}_i(\psi).\end{aligned}$$

Proof of Unbiasedness for the Overall Effect Estimate (based on Hudgens and Halloran [2008]): Again we'll show the unbiasedness result for $\hat{Y}(1|\psi)$ as the proof naturally extends to $\hat{Y}(0|\psi)$, $\hat{Y}(1|\phi)$ and $\hat{Y}(0|\phi)$ in the same fashion. We use the fact that $\mathbb{E}\left(\hat{Y}(1|\psi)\right) = \mathbb{E}\left(\mathbb{E}\left(\hat{Y}(1|\psi)|\mathbf{S}\right)\right)$, and from the fact that $\mathbb{E}[\hat{Y}_i(1|\psi)] = \bar{Y}_i(1|\psi)$, we have that $\mathbb{E}\left(\hat{Y}(1|\psi)\right) = \mathbb{E}\left(\sum_{i=1}^N \bar{Y}_i(1|\psi)S_i/C\right) = \bar{Y}(1|\psi)$. Recall that $C = \sum_i S_i$, the number of groups assigned to treatment strategy ψ.

5.2.2 Inverse-Probability-Weighted Estimators for Bernoulli Randomization

Inverse probability-weighted estimators for each type of effect are defined assuming ϕ and ψ are Bernoulli randomization strategies.

Under *Assumptions 3 and 4*, i.e., assuming that the treatment assignment mechanism satisfies the group-level conditional randomization and positivity assumptions, and that ψ is the parameterization of a Bernoulli (type B) strategy, then let

$$\hat{Y}_i^{ipw}(a|\psi) \equiv \frac{\sum_{j=1}^{n_i} Pr_\psi(\mathbf{A}_{i(j)} = \mathbf{A}_{i(j)}|L_i) I(A_{ij} = a)\mathbf{Y}_{ij}(\mathbf{A}_i)}{n_i Pr_\psi(A_i = 1|L_i)}, \qquad \text{and}$$

$$\hat{Y}_i^{ipw}(\psi) \equiv \frac{\sum_{j=1}^{n_i} Pr_\psi(\mathbf{A}_{i(j)} = \mathbf{A}_{i(j)}|L_i)\mathbf{Y}_{ij}(\mathbf{A}_i)}{n_i Pr_\psi(A_i = 1|L_i)}.$$

Then $\mathbb{E}[\hat{Y}_i^{ipw}(a|\psi)] = \bar{Y}_i(a|\psi)$, and $\mathbb{E}[\hat{Y}_i^{ipw}(\psi)] = \bar{Y}_i(\psi)$.

Proof of Unbiasedness of the IPW Estimators (based on Tchetgen and VanderWeele [2012]): Recall that the IPW estimators are derived under assumptions 3 and 4, and under Bernoulli (type B) parameterization. Given the Bernoulli

randomized nature of the type B parameterization, for the simplicity of the proof, let p_ψ denote the constant expected proportion of treated in the ψ strategy arm. In our simulation study example, $\alpha_\psi = 50\%$, for example.

Under assumptions 3 and 4, we have that

$$\begin{aligned}
\mathbb{E}\left(\hat{Y}_i(a|\psi)\right) &= \sum_{\omega \in \psi^{n_i}} \left(\frac{\sum_{j=1}^{n_i} Pr_\psi(\mathbf{A}_{i(j)} = \omega_{(j)}|L_i) I(\omega_j = a) \mathbf{Y}_{ij}(\omega)}{n_i Pr_\psi(A_i = \omega|L_i)} \right) \cdot \prod_{j'=1}^{n} p_\psi^{\omega_{j'}} (1-p_\psi)^{1-\omega_{j'}} \\
&= \frac{1}{n_i} \sum_{\omega \in \psi^{n_i}} \frac{Pr_\psi(\mathbf{A}_i = \omega|L_i)}{Pr_\psi(\mathbf{A}_i = \omega|L_i)} \sum_{j=1}^{n_i} I(\omega_j = a) Y_{ij}(A_{ij} = \omega_j, \mathbf{A}_{i(j)} = \omega_{(j)}) \prod_{j' \neq j}^{n} p_\psi^{\omega_{j'}} (1-p_\psi)^{1-\omega_{j'}} \\
&= \frac{1}{n_i} \sum_{j=1}^{n_i} \sum_{\omega \in \psi^{n_i - 1}} Y_{ij}(A_{ij} = a, A_{i(j)} = \omega) \prod_{j' \neq j}^{n} p_\psi^{\omega_{j'}} (1-p_\psi)^{1-\omega_{j'}} \\
&= \bar{Y}_i(a|\psi),
\end{aligned}$$

where ψ^{n_i} represents the set of all possible treatment assignments to n_i units under randomization strategy ψ. Similarly, we have that

$$\begin{aligned}
\mathbb{E}[\hat{Y}_i^{ipw}(\psi)] &= \frac{1}{n_i} \sum_{\omega \in \psi^{n_i}} \frac{Pr_\psi(\mathbf{A}_i = \omega|\mathbf{L}_i)}{Pr_\psi(\mathbf{A}_i = \omega|L_i)} \sum_{j=1}^{n_i} Y_{ij}(\mathbf{A}_i = \omega) \prod_{j' \neq j}^{n} p_\psi^{\omega_{j'}} (1-p_\psi)^{1-\omega_{j'}} \\
&= \frac{1}{n_i} \sum_{j=1}^{n_i} \sum_{\omega \in \psi^{n_i}} Y_{ij}(\mathbf{A}_i = \omega) \prod_{j' \neq j}^{n} p_\psi^{\omega_{j'}} (1-p_\psi)^{1-\omega_{j'}} \\
&= \bar{Y}_i(\psi).
\end{aligned}$$

5.3 Variance Estimators and Confidence Intervals

In this section, we review the different variance estimators and methods for constructing confidence intervals that have been proposed in the literature. Along with proposing estimators for the various interference effects of interest, Hudgens and Halloran [2008] also propose variance estimators under assumption 2.

Assumption 2 states that an individual's counterfactual outcome depends only on their exposure level and on the total number of people who received the intervention in the same group. In other words, we assume that we can partition the set of all possible treatment assignments for group i into strata where the number of other people in the group who received the intervention is the same; moreover, in each of these strata, we assume there is no interference. Recall that without this assumption (when we make no assumptions about the interference in a group), when an individual's treatment is a binary random variable, for a given $A_{ij} = a$, there are $2^{n_i - 1}$ potential outcomes for an individual. Moreover, if we were to assume no interference, then each individual would only have one potential outcome associated with a given $A_{ij} = a$. Assumption 2 falls somewhere between these extremes, implying that an individual has n_i potential outcomes for a given $A_{ij} = a$. Importantly, without additional assumptions like Assumption 2, Hudgens and Halloran [2008] present an example affirming the notion that no unbiased estimators of $Var(\hat{Y}_i(a|\psi) \mid S_i = \psi)$ exist, because one can construct examples where the observed data is the same, but $Var(\hat{Y}_i(a|\psi) \mid S_i = \psi)$ will be different. Define $\hat{\sigma}^2_{ia}(\psi)$ as the within-group sample variance and $\hat{\sigma}^2_{ga}(\psi)$ as the between-group sample variance for individuals with $A_{ij} = a$:

$$\begin{aligned}
\hat{\sigma}^2_{ia}(\psi) &= \sum_{j=1}^{n_i} [Y_{ij}(a \mid \psi) - \hat{Y}_i(a|\psi)]^2 I(A_{ij} = a)/(K_{\psi,n_i} - 1) \\
&= \sum_{j=1}^{n_i} [Y_{ij} - \hat{Y}_i(a|\psi)]^2 I(A_{ij} = a)/(K_{\psi,n_i} - 1) \\
\hat{\sigma}^2_{ga}(\psi) &= \sum_{i=1}^{N} [\hat{Y}_i(a|\psi) - \hat{Y}(a|\psi)]^2 I(S_i = \psi)/(C-1)
\end{aligned}$$

where $C = \sum_{i=1}^{N} I(S_i = \psi)$ and $Y_{ij}(a \mid \psi)$ is defined as being equal to $Y_{ij}(A_{ij} = a, \mathbf{A}_{i(j)} = \boldsymbol{\omega})$ for any $\boldsymbol{\omega} \in \psi^{n_i - 1}$, which is equal for all ω from Assumption 2, and hence $Y_{ij}(a \mid \psi) I(A_{ij} = a) = Y_{ij} I(A_{ij} = a)$ from consistency.

Moreover, define:

$$\hat{\sigma}^2_{DE}(\psi) = \sum_{i=1}^{N} [\hat{DE}_i(\psi) - \hat{DE}(\psi)]^2 I(S_i = \psi)(C-1)$$
$$\hat{\sigma}^2_{M}(\psi) = \sum_{i=1}^{N} [\hat{Y}_i(a|\psi) - \hat{Y}(a|\psi)]^2 I(S_i = \psi)(C-1).$$

Hudgens and Halloran [2008] present the following estimators for the variance of our quantities of interest:

$$\hat{Var}\{\hat{DE}_i(\psi) \mid S_i = \psi\} = \frac{\hat{\sigma}^2_{i1}(\psi)}{K_{\psi,n_i}} + \frac{\hat{\sigma}^2_{i0}(\psi)}{n_i - K_{\psi,n_i}}$$
$$\hat{Var}\{\hat{DE}(\psi) \mid S_i = \psi\} = (1 - \frac{C}{N})\hat{\sigma}^2_{DE}(\psi) + \frac{1}{CN}\sum_{i=1}^{N} \hat{Var}\{\hat{DE}_i(\psi) \mid S_i = \psi\} I(S_i = \psi)$$
$$\hat{Var}\{\hat{IE}(\psi,\phi)\} = \frac{\hat{\sigma}^2_{g0}(\psi)}{N-C} + \frac{\hat{\sigma}^2_{g0}(\phi)}{C}$$
$$\hat{Var}\{\hat{TE}(\psi,\phi)\} = \frac{\hat{\sigma}^2_{g0}(\psi)}{N-C} + \frac{\hat{\sigma}^2_{g1}(\phi)}{C}$$
$$\hat{Var}\{\hat{OE}(\psi,\phi)\} = \frac{\hat{\sigma}^2_{M}(\psi)}{N-C} + \frac{\hat{\sigma}^2_{M}(\phi)}{C}$$

Under Assumptions 1 and 2, Hudgens and Halloran [2008] show that these aforementioned estimators are generally biased upward without additional assumptions. For instance, if there were a constant additive individual direct effect across all groups such that $Y_{ij}(1 \mid \psi) = Y_{ij}(0 \mid \psi) + \psi_{D,i}$, then the proposed variance estimator for the direct effect would be unbiased. Hudgens and Halloran [2008] present analogous assumptions for the estimators of the indirect, total, and overall effects to be unbiased.

Tchetgen and VanderWeele [2012] construct finite sample exact confidence intervals for these population causal effects of interest, focusing on the binary outcome setting. Using a Hoeffding-type exponential inequality, they propose the following conservative $(1-\alpha)$ confidence intervals, which are guaranteed under Assumption 1 to have coverage probability no smaller than $1-\alpha$:

$$CI_{DE}(\alpha) = (\hat{DE}(\psi) - \epsilon^*_{DE}(\alpha), \hat{DE}(\psi) + \epsilon^*_{DE}(\alpha))$$
$$CI_{IE}(\alpha) = (\hat{IE}(\psi) - \epsilon^*_{IE}(\alpha), \hat{IE}(\psi) + \epsilon^*_{IE}(\alpha))$$
$$CI_{TE}(\alpha) = (\hat{TE}(\psi) - \epsilon^*_{TE}(\alpha), \hat{TE}(\psi) + \epsilon^*_{TE}(\alpha))$$
$$CI_{OE}(\alpha) = (\hat{OE}(\psi) - \epsilon^*_{OE}(\alpha), \hat{OE}(\psi) + \epsilon^*_{OE}(\alpha))$$

where for the direct effect, we have

$$\epsilon^*_{DE}(\alpha) = \sqrt{(2N)^{-1}\left[4(\frac{N}{C} - 1)^2 + N^{-1}\sum_{i=1}^{N} 4L^2_{\psi,i}\left(\frac{N}{C}\right)^2\right]\ln\left(\frac{2}{\alpha}\right)}$$

and

$$L_{\psi,i} = \left(1 - \binom{n_i}{K_{\psi,n_i}}^{-1}\right), \qquad L_{\phi,i} = \left(1 - \binom{n_i}{K_{\phi,n_i}}^{-1}\right).$$

For the indirect, total, and overall effects, we have that

$$\epsilon^*_{TE}(\alpha) = \epsilon^*_{OE}(\alpha) = \epsilon^*_{IE}(\alpha)$$
$$= \sqrt{(2N)^{-1}\left[\max\left\{\frac{1}{(C/N)^2}, \frac{1}{(1-C/N)^2}\right\} + N^{-1}\sum_{i=1}^{N}\max\left\{\frac{1}{C/N}L_{\psi,i}, \frac{1}{1-C/N}L_{\phi,i}\right\}^2\right]\ln\left(\frac{2}{\alpha}\right)}.$$

Tchetgen and VanderWeele [2012] note that we can similarly construct intervals for the population conditional average causal direct effect, that is the average causal direct effect for groups randomized to a given treatment strategy. When

such a quantity is of interest, these intervals will be centered around the same estimator for the direct effect and will generally be significantly more precise.

Liu and Hudgens [2014] propose a method for constructing confidence intervals based on the asymptotic distribution of the estimators mentioned in the previous section. They consider asymptotics under two regimes: one where the size of each group n_i grows large (but where the number of groups N is not required to) and the other where the number of groups N grows large (but the size of each group n_i is not required to). Under the former asymptotic regime (along with some additional assumptions), the population direct, indirect, total, and overall effect estimators asymptotically are distributed as a mixture of Normal random variables. When there is an additional assumption of mean homogeneity, i.e., the mean causal effect is the same across all groups, then we have that the asymptotic distribution simply follows a Normal distribution. Similarly, under the latter asymptotic regime, relying on the Lindeberg condition, we have that the distribution of the causal effect estimators also follows a Normal distribution, even without assuming mean homogeneity across groups. Lastly, to construct approximate large sample confidence intervals, Liu and Hudgens [2014] construct standard Wald CIs when the limiting distribution is Normal and use Chebyshev's inequality to construct conservative CIs when the limiting distribution is a mixture of normals (i.e., under the asymptotic regime where group size is large and mean homogeneity is not assumed across groups).

Rigdon and Hudgens [2015] propose a method for exact confidence intervals of these population interference effects by inverting permutation tests when the outcome is binary and Assumption 2 holds. Their general procedure, referred to as the exact inverted test (EIT) is summarized below for $DE(\psi)$, although the approach is easily generalizable to other inference quantities of interest:

1. **Enumerate all sharp null hypotheses consistent with the observed data.** Under Assumption 2, each individual has 2 potential outcomes relevant for $DE(\psi)$: $Y_{ij}(1,\psi)$ and $Y_{ij}(1,\phi)$. For individuals in groups assigned to ψ, one potential outcome is observed ($Y_{ij}(a,\psi)$), whereas for individuals in groups assigned to ϕ, neither potential outcome is observed. Permuting over possible values of the unobserved potential outcomes enumerates all sharp null hypotheses consistent with the observed data. Note that each null hypothesis will imply a corresponding null value for the direct effect, $DE^0(\psi)$.
2. **Test each sharp null hypothesis using a permutation test.** For a given sharp null hypothesis, an exact p-value for the two-sided test can be computed by permuting over all group and treatment assignments consistent with the two-stage randomization strategy used to generate the observed data (i.e., keeping C, ψ, and ϕ the same). From each group and treatment assignment permutation, one can calculate the corresponding $\hat{DE}_c(\psi)$. The p-value $p(DE^0(\psi))$ is then defined as the proportion of permutations where the absolute deviation between the $\hat{DE}_c(\psi)$ and $DE^0(\psi)$ is larger than that between the estimator $\hat{DE}(\psi)$ and $DE^0(\psi)$. When multiple null hypothesis correspond to the same value of $DE^0(\psi)$, define $p(DE^0(\psi))$ to be the maximum p-value associated with testing all such hypotheses.
3. **Form the confidence interval by inverting the hypothesis tests.** A $1-\alpha$ confidence set for $DE(\psi)$ can be defined as $CI_{DE}(\alpha) = \{DE^0(\psi) : p(DE^0(\psi)) \geqslant \alpha\}$.

Rigdon and Hudgens [2015] then propose a computationally feasible algorithm to approximate the confidence sets, given the potentially large number of operations needed to run a permutation test for potentially many null hypotheses. They also carry out a simulation study, in which they present potential benefits of the EIT confidence interval over those previously derived in Tchetgen and VanderWeele [2012] and Liu and Hudgens [2014].

6 Simulation Study

For all simulation studies we assume a randomized trial with no covariates. We assume the outcome is linear model with a term for an individual's own treatment as well as an "interference" term for each other individual it's group. The interference terms are prefixed by a parameter γ that controls the level of interference, where $\gamma = 0$ implies no interference.

$$Y_{ij}(\mathbf{a}) = \beta_0 + \beta_1 a_j + \gamma \sum_{k \neq j} \beta_2 a_k + \epsilon_{ij}$$
$$\epsilon_{ij} \sim N(0, \sigma^2 = 0.1)$$

We let the baseline expected level of individual infection when everyone is untreated is $\beta_0 = 10$. We assume an individual experiences a 5 point reduction in their level of infection due to their own treatment with $\beta_1 = -5$. Further, we assume an individual experiences a 0.1 point reduction in level of infection due to the treatment of each other individual with $\beta_2 = -0.1$.

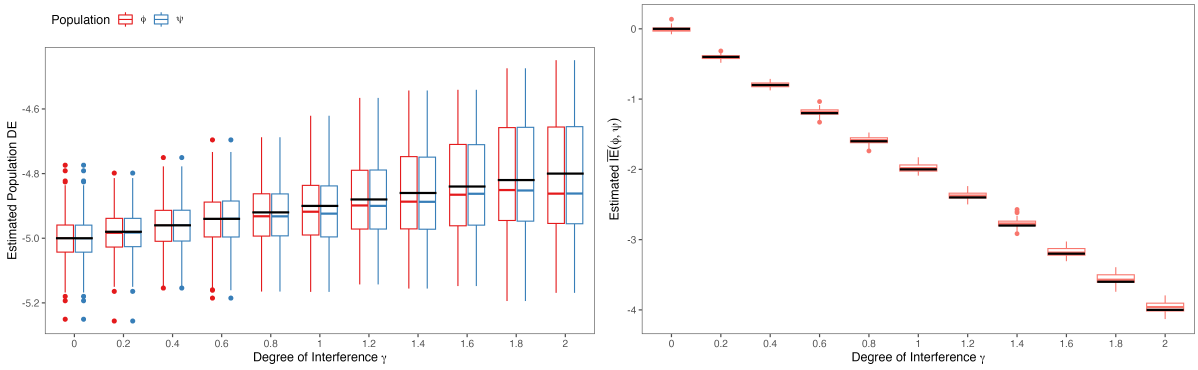

Figure 4: **Finite Sample Setting**: Estimated population average direct (left) and indirect (right) effects for assignment strategies ϕ, where 50 of 100 individuals are treated, and ψ, where 30 of 100 individuals are treated. True values are marked as horizontal black lines: $\overline{DE}(\phi) = \overline{DE}(\psi) = \beta_1 - \gamma\beta_2 = -5 + 0.1\gamma$ and $\overline{IE}(\phi, \psi) = \gamma\beta_2(K_{\phi,n_i} - K_\psi) = -2\gamma$. For each value of γ, 100 datasets were simulated.

6.1 Randomization Type A: Finite Sample & Standard Direct Effect

In this section, we compute the true population average direct and indirect effects analytically under a finite sample setting (randomization type A). We compare these analytical ground truths to estimates on simulated data.

We implement the finite sample randomization procedure as follows: first, the sampling procedure S_i is chosen for each group i. Then, treatments A_{ij} are assigned according to the assigned sampling procedure.

$$S_i \sim \text{finite sample from } \{\psi, \psi, \psi, \phi, \phi\}$$

$$A_{ij} \sim \begin{cases} \text{finite sample from 30 treated, 70 untreated, without replacement} & \text{if } S_i = \psi \\ \text{finite sample from 50 treated, 50 untreated, without replacement} & \text{if } S_i = \phi \end{cases}$$

Section 3.1 defines the standard formulation of the population average direct effect as $\overline{DE}(\phi) = \bar{Y}(1|\phi) - \bar{Y}(0|\phi) = \frac{1}{N}\sum_{i=1}^{N} \overline{DE}_i(\phi)$. Below, we compute analytically what this true direct effect is. Define K_{ϕ,n_i} as the fixed number of individuals treated under assignment strategy ϕ versus K_ψ as the number under assignment strategy ψ.

$$\begin{aligned}
\bar{Y}_{ij}(a|\phi) &= \sum_{\mathbf{s}\in\phi^{n_i-1}} Y_{ij}(\mathbf{a}_{i(j)} = \mathbf{s}, a_{ij} = a) \cdot Pr_\phi(\mathbf{A}_{i(j)} = \mathbf{s}|A_{ij} = a) \\
&= \sum_{\mathbf{s}\in\phi^{n_i-1}} \left(\beta_0 + \beta_1 a + \gamma\beta_2 \sum_{k\neq j} s_k + \epsilon_{ij}\right) \cdot \frac{1}{\binom{n-1}{K_{\phi,n_i}-a}} \cdot I\Big(\sum_{k\neq j} s_k = K_{\phi,n_i} - a\Big) \\
&= (\beta_0 + \beta_1 a_j + \gamma\beta_2(K_{\phi,n_i} - a) + \epsilon_{ij}) \cdot \frac{\binom{n-1}{K_{\phi,n_i}-a}}{\binom{n-1}{K_{\phi,n_i}-a}} \\
&= \beta_0 + \beta_1 a + \gamma\beta_2(K_{\phi,n_i} - a) + \epsilon_{ij} \\
\bar{Y}_i(a|\phi) &= \frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ij}(a|\phi) = \beta_0 + \beta_1 a + \gamma\beta_2(K_{\phi,n_i} - a) + \frac{1}{n_i}\sum_{j=1}^{n_i}\epsilon_{ij} \\
\overline{DE}_i(\phi) &= \bar{Y}_i(1|\phi) - \bar{Y}_i(0|\phi) = \beta_1 - \gamma\beta_2 \\
\overline{DE}(\phi) &= \frac{1}{N}\sum_{i=1}^{N}\overline{DE}_i(\phi) = \beta_1 - \gamma\beta_2
\end{aligned}$$

We see that the true population average direct effect does not depend on K_{ϕ,n_i}, so will be identical for assignment strategy ψ. Further, we perhaps unexpectedly see that this "direct effect" depends on the degree of interference γ and the interference coefficient β_2. This is the problem identified in VanderWeele and Tchetgen Tchetgen [2011] that will be remedied in the following section using Bernoulli Randomization (type B).

Section 3.2 defines the standard formulation of the population average indirect effect as $\overline{IE}(\phi, \psi) = \bar{Y}(0|\phi) - \bar{Y}(0|\psi)$. We compute analytically what this true direct effect is under a finite sample setting and our defined data generating function.

$$\bar{Y}(a|\phi) = \frac{1}{N}\sum_{i=1}^{N} \bar{Y}_i(a|\phi) = \beta_0 + \beta_1 a + \gamma\beta_2(K_{\phi,n_i} - a) + \frac{1}{N}\sum_{i=1}^{N}\frac{1}{n_i}\sum_{j=1}^{n-i}\epsilon_{ij}$$
$$\overline{IE}(\phi, \psi) = \bar{Y}(0|\phi) - \bar{Y}(0|\psi) = \gamma\beta_2(K_{\phi,n_i} - K_\psi)$$

In our simulation studies, we use the standard estimators from 3 based on $\hat{Y}_i(a|\phi) = \frac{\sum_{j=1}^{n_i} I(A_{ij}=a)Y_{ij}(\mathbf{A}_i)}{\sum_{j=1}^{n_i} I(A_{ij}=a)}$ [Hudgens and Halloran, 2008]. Figure 4 (left) shows estimated population average direct effects across 100 simulated datasets for each value of degree of interference γ. Estimates are centered at our empirically derived true population average direct effect of $\beta_1 - \gamma\beta_2 = -5 + 0.1\gamma$. We also see that the variability of these estimates increase with the degree of interference.

We see that the true population average indirect effect only depends on the difference between the assignment strategies, as one would expect. Figure 4 (right) shows estimated values across 100 datasets for each value of degree of interference γ. Estimates are centered at the true population average indirect effect $\gamma\beta_2(K_{\phi,n_i} - K_\psi) = -2\gamma$. This negative value makes sense: we expect a protective effect of vaccinating 50 individuals versus only 30.

6.2 Randomization Type B: Bernoulli & Redefined Direct Effect

In this section, we consider a Bernoulli randomization setting (randomization type B). Importantly, the redefined direct effect version from VanderWeele and Tchetgen Tchetgen [2011] which results in a more intuitive "direct effect" that does not depend on interference at all. We analytically derive the true value of this new direct effect and compare to estimates on simulated data.

We implement the Bernoulli randomization procedure identically to the fixed sample randomization procedure, but this time A_{ij}s are sampled independently rather than from a finite population.

$$S_i \sim \text{finite sample from } \{\psi, \psi, \psi, \phi, \phi\}$$
$$A_{ij} \sim \begin{cases} \text{Bernoulli}(0.3) & \text{if } S_i = \psi \\ \text{Bernoulli}(0.5) & \text{if } S_i = \phi \end{cases}$$

Define p_ϕ as the probability of being assigned treatment under assignment strategy ϕ.

$$\begin{aligned}
\bar{Y}^*_{ij}(a|\phi, a') &= \sum_{\omega\in\psi^{n_i-1}} Y_{ij}(\mathbf{a}_{i(j)} = \omega, a_{ij} = a) Pr_\psi(\mathbf{A}_i = \omega | a_{ij} = a') \\
&= \sum_{\omega\in\psi^{n_i-1}} \left(\beta_0 + \beta_1 a + \gamma\beta_2 \sum_{k\neq j}\omega_k + \epsilon_{ij}\right) \cdot p_\phi^{\sum\omega}(1-p_\phi)^{n_i - \sum\omega} \cdot I(\omega_j = a') \\
\overline{DE}^*_{ij}(\psi, a) &= \bar{Y}^*_{ij}(1|\psi, a) - \bar{Y}^*_{ij}(0|\psi, a) = \beta_1 \\
\overline{DE}^*(\psi) &= \frac{1}{N}\sum_{i=1}^{N}\frac{1}{n_i}\sum_{j=1}^{n_i}\overline{DE}^*_{ij}(\psi, a) = \beta_1
\end{aligned}$$

We see that the alternate definition of the true population average direct effect does not depend on the degree of interference γ or the interference coefficient β_2. This makes more sense as a "direct effect" because it remains constant regardless of interference.

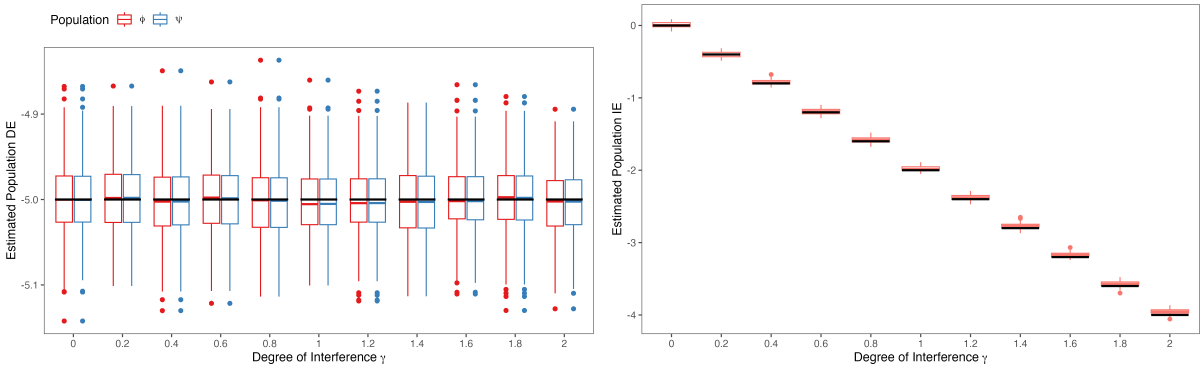

Figure 5: **Bernoulli Sample Setting**: Estimated population average direct (left) and indirect effects (right) effects for assignment strategies ϕ, where 50% of individuals are treated, and ψ, where 30% of individuals are treated. True values are marked as horizontal black lines: $\overline{DE}^*(\phi) = \overline{DE}^*(\psi) = \beta_1 = -5$ and $\overline{IE}^*(\phi, \psi) = \gamma n \beta_2 (p_\phi - p_\psi) = -2\gamma$. For each value of γ, 100 datasets were simulated.

$$
\begin{aligned}
\bar{Y}^*_{ij}(a|\phi) &= \sum_{\omega \in \phi^{n_i}} Y_{ij}(\mathbf{a}_{i(j)} = \omega_{(j)}, a_{ij} = a) Pr_\phi(\mathbf{A}_i = \omega) \\
&= \sum_{\omega \in \phi^{n_i}} \left(\beta_0 + \beta_1 a + \gamma \beta_2 \sum_{k \neq j} \omega_k + \epsilon_{ij} \right) \cdot p_\phi^{\sum \omega} (1 - p_\phi)^{n_i - \sum \omega} \\
&= \sum_{t=0}^{n} (\beta_0 + \beta_1 a + \gamma \beta_2 t + \epsilon_{ij}) \cdot p_\phi^t (1 - p_\phi)^{n_i - t} \text{ where } t = \sum \omega \\
\overline{IE}^*_{ij}(\phi, \psi) &= \bar{Y}^*_{ij}(0|\phi) - \bar{Y}^*_{ij}(0|\psi) \\
&= \sum_{t=0}^{n} (\beta_0 + \beta_1 + \gamma \beta_2 t + \epsilon_{ij}) \cdot p_\phi^t (1 - p_\phi)^{n_i - t} - \sum_{t=0}^{n} (\beta_0 + \gamma \beta_2 t + \epsilon_{ij}) \cdot p_\phi^t (1 - p_\phi)^{n_i - t} \\
&= \gamma \beta_2 \cdot (E_{t_\phi}[t] - E_{t_\psi}[t]) \text{ where } t_\phi \sim \text{Binomial}(n_i, p_\phi) \text{ and } t_\psi \sim \text{Binomial}(n_i, p_\psi) \\
&= \gamma n_i \beta_2 (p_\phi - p_\psi) \\
\overline{IE}^*(\phi, \psi) &= \frac{1}{N} \sum_{i=1}^{N} \frac{1}{n_i} \sum_{j=1}^{n_i} \overline{IE}^*_{ij}(\phi, \psi) = \gamma n \beta_2 (p_\phi - p_\psi) \text{ assuming } n_i = n \forall i
\end{aligned}
$$

The indirect effect is the same as before because $K_{\phi, n_i} = n p_\phi$ and $K_\psi = n p_\psi$.

In our simulation studies, we use the estimators for this alternative form of direct and indirect effects from VanderWeele and Tchetgen Tchetgen [2011] that are based on $\hat{Y}_i(a|\phi) = \frac{\sum_{j=1}^{n_i} I(A_{ij}=a) Y_{ij}(\mathbf{A}_i)}{\sum_{j=1}^{n_i} I(A_{ij}=a)}$, which align with the finite sample estimators from Hudgens and Halloran [2008]. Figure 5 (left) shows that the simulation studies estimate this value very well. Further we see that the variability of these estimates is stable as degree of interference increases. Figure 5 (right) mirrors Figure 4 (right) because the indirect effect is the same under the original and new formulations.

7 Applications

In this section we review a real-data application of the concepts discussed in this paper.

We consider the following motivating example from VanderWeele and Tchetgen [2011], originally reported in Millar et al. [2008]. In a group-randomized vaccine trial of a 7-valent pneumococcal conjugate vaccine (PCV7) among southwestern American Indian communities, households were randomized to either have one child receive the control vaccine (a meningococcal conjugate vaccine against serogroup C [MCC]) or to have one child receive PCV7. Here, the "infectious" effects of interest are of the vaccine — the child being vaccinated has a protective effect against other children and adults in the household against pneumococcal colonization. The authors found that adults and unvaccinated

children aged <5 years living in households with a PCV7 vaccinated child were less likely to be colonized than those living in a household with a control vaccinated child. The authors conclude that "vaccine-type pneumococcal carriage was lower among adults and unvaccinated children living with a PCV7 vaccinee. This was attributable to reduced exposure and reduced transmission when exposure occurs."

8 Limitations

We present a potentially non-exhaustive list of limitations of the methods for adjusting for interference presented above:

- **Standard assumptions apply**. As is usual, we need a version of positivity and consistency (though this time at the group-level), and conditional randomization to hold.
- **Stratified and Partial Interference**. The methods presented strongly rely on the assumptions of partial interference (which refers to the assumption that the interference effect is only present within identified groups) and stratified interference (which refers to the condition that the interference effect is only through the proportion of the group that is treated). These are strong assumptions, but they buy us a lot in terms of identifiability, though they are not always realistic.
- **Group-Randomized Design**. A direct implication of the partial interference assumption is that our data resemble or are resultant from a *group-randomized design*. Some examples of studies following this design can be found in the literature as 'parallel groups' or 'parallel-group randomized trials' Millar et al. [2008], Murray [1998], Small et al. [2008]. Anecdotally, while we could find examples of studies that were designed from the beginning as group-randomized trials, they were not particularly common and were challenging to find due to both variation in terminology and the similarity in language between 'group randomized' and 'cluster randomized' trials.
- **Contamination**. Though these methods do construct unbiased estimators when interference is limited to within groups, if there is interference across groups that is either unknown or ignored and these estimators are applied as if such contamination does not exist, they will be biased.
- **Assumption of Homogeneous Mixture**. One assumption that is repeatedly made throughout the works we looked at on interference in randomized (especially the group-randomized) settings is that what matters is the prevalence of treatment in an individuals' group. This implicitly assumes that the level of interference from other units are is equally weighted. In the example of a vaccine trial where groups are households and only some individuals are vaccinated, a violation of this assumption that we can imagine is if the interference effects were stronger for individuals in households with closer contact.
- **Covariate Shift**. As a particular case of the standard assumptions mentioned above, covariate shift refers to the situation where the distribution of covariates differs between the treated and untreated groups in ways that can result in positivity violations or in residual confounding between the treatment assignment mechanism and outcome.

9 Recent Works and Current Areas of Research

We consider for the purposes of search criteria those works that were published in 2020 or after on spillover effects in randomized trials. The following list is not meant to be exhaustive, but nonetheless intended to highlight a few works to give a flavor of topics in the literature.

- Aronow et al. [2021] in their book chapter *Spillover Effects in Experimental Data* (in *Advances in Experimental Political Science*) outline a number of motivating examples (anti-bullying school programs, unconditional cash-transfer programs, and get-out-the-vote efforts), present a Horvitz-Thompson style estimator, the `interference` R package for dealing with network intervention, and present simulations showing that when the interference dependency structure is misspecified, it induces bias.
- Anaya-Izquierdo and Alexander [2021] present a method for handling spillover effects in cluster randomized trials and illustrate the method using pair-matched cluster randomized trial data against the dengue mosquito vector *Aedes aegypti*.
- Sävje et al. [2021] investigate the large-sample properties of estimators under unknown levels of interference in randomized experiments, deriving rates of convergence.
- Buchanan et al. [2022] estimate the effects of an HIV intervention among people who inject drugs using marginal structural models with the `interference` package.

- Jiang et al. [2022] describe a framework for conducting power analysis in two-stage randomized experiments.
- Vazquez-Bare [2023a] analyzes the identification of causal direct and spillover effects under one-sided noncompliance and show that causal effects can be estimated by 2SLS as in instrumental variables.
- Brown et al. [2023] present a systematic review of spillover effects in childhood obesity prevention interventions.
- Vazquez-Bare [2023b] proposes nonparametric estimators for average direct and spillover effects that are consistent, asymptotically normal, and illustrated using data from a conditional cash transfer program and via simulation.
- Hernández-Ramírez et al. [2024] report on a peer-education intervention among people who inject drugs through an analysis of 226 egocentric groups/networks randomized to receive treatment or not, in particular, presenting overall, direct, spillover, and composite effects.

10 Discussion

In our work, we introduced and motivated the problem of estimating direct and spillover effects in randomized trials. Through simulation, we demonstrated how conflating the traditional average treatment effect estimator and a direct effect estimator that takes interference into account would lead to inferential error. While total and overall effects can be validly estimated without taking interference into account and may often be of scientific interest, the disaggregation of total effects into direct effects and those due to spillover continues to remain relevant in fields like vaccine trials, epidemiology, and economics.

The treatment of spillover effects we presented requires the introduction of characterizations of fixed sample (type A, fixed K_{ψ,n_i} to be treated in each study group) and Bernoulli (type B, a fixed probability of treatment p_ψ) parameterized randomization strategies. Doing so allows us to formalize total effects and show how they are decomposed into direct and spillover (indirect) effects, as well as the overall effect. After presenting estimators and their derivations, we continue on to report on a simulation study. Figure 1 shows the type of mistaken thinking one could engage in if one simply adopted the ATE in the context of spillover effects, and Figure 4 demonstrates our empirical estimation of population direct and indirect effects under varying degrees of interference. Finally, we offer several examples from the current literature to illustrate how this body of theory is being used in practice.

Though the estimation of spillover effects and effects due to interference in randomized studies is a relatively new area of research, it is a significant one that has the potential to inform policy and practice in a variety of fields. We hope that this manuscript has provided a clear and accessible introduction to the estimation of causal effects in the presence of partial interference.

References

David Roxbee Cox. The interpretation of the effects of non-additivity in the latin square. *Biometrika*, 45(1/2):69–73, 1958.

Donald B Rubin. Randomization analysis of experimental data: The fisher randomization test comment. *Journal of the American statistical association*, 75(371):591–593, 1980.

Michael G Hudgens and M. Elizabeth Halloran. Toward Causal Inference With Interference. *Journal of the American Statistical Association*, 103(482):832–842, June 2008. ISSN 0162-1459, 1537-274X. doi:10.1198/016214508000000292. URL https://www.tandfonline.com/doi/full/10.1198/016214508000000292.

Michael E Sobel. What Do Randomized Studies of Housing Mobility Demonstrate?: Causal Inference in the Face of Interference. *Journal of the American Statistical Association*, 101(476):1398–1407, December 2006. ISSN 0162-1459, 1537-274X. doi:10.1198/016214506000000636. URL https://www.tandfonline.com/doi/full/10.1198/016214506000000636.

Ronald Ross. An application of the theory of probabilities to the study of a priori pathometry.—part i. *Proceedings of the Royal Society of London. Series A, Containing papers of a mathematical and physical character*, 92(638): 204–230, 1916.

M.Elizabeth Halloran, Michael Haber, Ira M. Longini, and Claudio J. Struchiner. Direct and Indirect Effects in Vaccine Efficacy and Effectiveness. *American Journal of Epidemiology*, 133(4):323–331, February 1991. ISSN 1476-6256, 0002-9262. doi:10.1093/oxfordjournals.aje.a115884. URL https://academic.oup.com/aje/article/166589/Direct.

Claudio J Struchiner, Mary Elizabeth Halloran, James M Robins, and Andrew Spielman. The Behaviour of Common Measures of Association Used to Assess a Vaccination Programme under Complex Disease Transmission Patterns—A Computer Simulation Study of Malaria Vaccines. *International Journal of Epidemiology*, 19(1):187–196, 1990. ISSN 0300-5771, 1464-3685. doi:10.1093/ije/19.1.187. URL https://academic.oup.com/ije/article-lookup/doi/10.1093/ije/19.1.187.

M. Elizabeth Halloran and Claudio J. Struchiner. Causal Inference in Infectious Diseases:. *Epidemiology*, 6(2):142–151, March 1995. ISSN 1044-3983. doi:10.1097/00001648-199503000-00010. URL http://journals.lww.com/00001648-199503000-00010.

Eric J Tchetgen Tchetgen and Tyler J VanderWeele. On causal inference in the presence of interference. *Statistical Methods in Medical Research*, 21(1):55–75, February 2012. ISSN 0962-2802, 1477-0334. doi:10.1177/0962280210386779. URL http://journals.sagepub.com/doi/10.1177/0962280210386779.

Paul R Rosenbaum. Interference Between Units in Randomized Experiments. *Journal of the American Statistical Association*, 102(477):191–200, March 2007. ISSN 0162-1459, 1537-274X. doi:10.1198/016214506000001112. URL http://www.tandfonline.com/doi/abs/10.1198/016214506000001112.

M. Elizabeth Halloran and Michael G. Hudgens. Dependent Happenings: a Recent Methodological Review. *Current Epidemiology Reports*, 3(4):297–305, December 2016. ISSN 2196-2995. doi:10.1007/s40471-016-0086-4. URL http://link.springer.com/10.1007/s40471-016-0086-4.

Louis Dron, Monica Taljaard, Yin Bun Cheung, Rebecca Grais, Nathan Ford, Kristian Thorlund, Fyezah Jehan, Etheldreda Nakimuli-Mpungu, Denis Xavier, Zulfiqar A Bhutta, et al. The role and challenges of cluster randomised trials for global health. *The Lancet Global Health*, 9(5):e701–e710, 2021.

Alejandra Benitez, Maya L Petersen, Mark J van der Laan, Nicole Santos, Elizabeth Butrick, Dilys Walker, Rakesh Ghosh, Phelgona Otieno, Peter Waiswa, and Laura B Balzer. Defining and estimating effects in cluster randomized trials: A methods comparison. *Statistics in medicine*, 42(19):3443–3466, 2023.

Jade Benjamin-Chung, Benjamin F Arnold, David Berger, Stephen P Luby, Edward Miguel, John M Colford Jr, and Alan E Hubbard. Spillover effects in epidemiology: parameters, study designs and methodological considerations. *International Journal of Epidemiology*, pages 332–347, 2018. doi:10.1093/ije/dyx201. URL https://academic.oup.com/ije/article/doi/10.1093/ije/dyx201/4567296. Advance Access Publication Date: 2 November 2017.

Tyler J. VanderWeele and Eric J. Tchetgen Tchetgen. Effect partitioning under interference in two-stage randomized vaccine trials. *Statistics & Probability Letters*, 81(7):861–869, July 2011. ISSN 01677152. doi:10.1016/j.spl.2011.02.019. URL https://linkinghub.elsevier.com/retrieve/pii/S0167715211000654.

Lan Liu and Michael G. Hudgens. Large Sample Randomization Inference of Causal Effects in the Presence of Interference. *Journal of the American Statistical Association*, 109(505):288–301, January 2014. ISSN 0162-1459, 1537-274X. doi:10.1080/01621459.2013.844698. URL http://www.tandfonline.com/doi/abs/10.1080/01621459.2013.844698.

Joseph Rigdon and Michael G. Hudgens. Exact confidence intervals in the presence of interference. *Statistics & Probability Letters*, 105:130–135, October 2015. ISSN 01677152. doi:10.1016/j.spl.2015.06.011. URL https://linkinghub.elsevier.com/retrieve/pii/S0167715215001972.

Tyler J. VanderWeele and Eric J. Tchetgen Tchetgen. Bounding the infectiousness effect in vaccine trials. *Epidemiology*, 22(5):686–693, 2011. doi:10.1097/EDE.0b013e31822708d5. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3792580/.

Eugene V. Millar, James P. Watt, Melinda A. Bronsdon, Jean Dallas, Raymond Reid, Mathuram Santosham, and Katherine L. O'Brien. Indirect effect of 7-valent pneumococcal conjugate vaccine on pneumococcal colonization among unvaccinated household members. *Clinical Infectious Diseases*, 47(8):989–996, 2008. doi:10.1086/591966. URL https://doi.org/10.1086/591966.

David M. Murray. *Design and Analysis of Group Randomized Trials*. Oxford University Press, New York, NY, 1998.

Dylan S Small, Thomas R Ten Have, and Paul R Rosenbaum. Randomization Inference in a Group–Randomized Trial of Treatments for Depression: Covariate Adjustment, Noncompliance, and Quantile Effects. *Journal of the American Statistical Association*, 103(481):271–279, March 2008. ISSN 0162-1459, 1537-274X. doi:10.1198/016214507000000897. URL https://www.tandfonline.com/doi/full/10.1198/016214507000000897.

Peter M. Aronow, Dean Eckles, Cyrus Samii, and Stephanie Zonszein. Spillover effects in experimental data. In James Druckman and Donald P. Green, editors, *Advances in Experimental Political Science*, pages 289–319. Cambridge University Press, 2021. ISBN 9781108478502.

Karim Anaya-Izquierdo and Neal Alexander. Spatial regression and spillover effects in cluster randomized trials with count outcomes. *Biometrics*, 77:490–505, 2021. doi:10.1111/biom.13316. URL https://doi.org/10.1111/biom.13316.

Fredrik Sävje, Peter M. Aronow, and Michael G. Hudgens. Average treatment effects in the presence of unknown interference. *The Annals of Statistics*, 49(2), April 2021. ISSN 0090-5364. doi:10.1214/20-AOS1973. URL https://projecteuclid.org/journals/annals-of-statistics/volume-49/issue-2/Average-treatment-effects-in-the-presence-of-unknown-interference/10.1214/20-AOS1973.full.

Ashley L. Buchanan, Raúl Ulises Hernández-Ramírez, Judith J. Lok, Sten H. Vermund, Samuel R. Friedman, Laura Forastiere, and Donna Spiegelman. Assessing direct and spillover effects of intervention packages in network-randomized studies. *medRxiv*, 2022. doi:10.1101/2022.03.24.22272909. URL https://doi.org/10.1101/2022.03.24.22272909.

Zhichao Jiang, Kosuke Imai, and Anup Malani. Statistical inference and power analysis for direct and spillover effects in two-stage randomized experiments. *Biometrics*, 2022. doi:10.1111/biom.13782. URL https://doi.org/10.1111/biom.13782.

Gonzalo Vazquez-Bare. Causal spillover effects using instrumental variables. *Journal of the American Statistical Association*, 118:1911–1922, 2023a. doi:10.1080/01621459.2021.2021920. URL https://doi.org/10.1080/01621459.2021.2021920. Published online: 03 February 2022.

Vicki Brown, Huong Tran, Jane Jacobs, Jaithri Ananthapavan, Claudia Strugnell, Kathryn Backholer, Marufa Sultana, Moosa Alsubhi, Steve Allender, Rachel Novotny, and Melanie Nichols. Spillover effects of childhood obesity prevention interventions: A systematic review. *Obesity Reviews*, 2023. doi:10.1111/obr.13692. URL https://doi.org/10.1111/obr.13692.

Gonzalo Vazquez-Bare. Identification and estimation of spillover effects in randomized experiments. *Journal of Econometrics*, 237:105237, 2023b. ISSN 0304-4076. doi:10.1016/j.jeconom.2021.10.014. URL https://www.sciencedirect.com/science/article/pii/S0304407621003067.

R.U. Hernández-Ramírez, D. Spiegelman, J.J. Lok, et al. Overall, direct, spillover, and composite effects of components of a peer-driven intervention package on injection risk behavior among people who inject drugs in the hptn 037 study. *AIDS Behavior*, 28:225–237, 2024. doi:10.1007/s10461-023-04213-x. URL https://doi.org/10.1007/s10461-023-04213-x.