Published in final edited form as:
Stat Methods Med Res. 2012 February ; 21(1): 55–75. doi:10.1177/0962280210386779.

On causal inference in the presence of interference

Eric J. Tchetgen Tchetgen and **Tyler J. VanderWeele**
Departments of Epidemiology and Biostatistics, Harvard University

Abstract

Interference is said to be present when the exposure or treatment received by one individual may affect the outcomes of other individuals. Such interference can arise in settings in which the outcomes of the various individuals come about through social interactions. When interference is present, causal inference is rendered considerably more complex, and the literature on causal inference in the presence of interference has just recently begun to develop. In this paper we summarize some of the concepts and results from the existing literature and extend that literature in considering new results for finite sample inference, new inverse probability weighting estimators in the presence of interference and new causal estimands of interest.

1 Introduction

Interference is said to be present when the exposure or treatment received by one individual may affect the outcomes of other individuals. Such interference can arise in settings in which the outcomes of the various individuals come about through social interactions (Manski, 2000, 2010). Most of the literature on causal inference proceeds by making an assumption of "no-interference." For example, Rubin's formulation of the potential outcomes framework an assumption referred to as the "Stable Unit Treatment Value Assumption" or "SUTVA" is made which includes within it a no-interference assumption (Rubin, 1980). Such no-interference assumptions are employed routinely though not always acknowledged. When interference is present, causal inference is rendered considerably more complex, and the literature on causal inference in the presence of interference has just recently begun to develop (Sobel, 2006; Hong and Raudenbush, 2006; Rosenbaum, 2007; Hudgens and Halloran, 2008; Graham, 2008; Manski, 2010). In this paper we hope to both summarize some of the concepts and results from the existing literature and to extend that literature in considering new results for finite sample inference, new inverse probability weighting estimators in the presence of interference and new causal estimands of interest.

The remainder of this paper is organized as follows. In section 2 we present the notation we will be using throughout. In section 3 we review notions of direct, indirect (spillover), total and overall causal effects of Hudgens and Halloran (2008) that arise when interference is present. In section 4 we discuss inference for these effects in randomized trials and present new results on variance estimation and finite sample confidence intervals in the presence of interference. In section 5 we consider the context of observational studies and present a result on inverse probability weighting estimators of causal effects when interference is present. In section 6, we discuss varieties of direct and indirect effects present in the causal inference literature and comment on the terminological ambiguity concerning the

expressions "direct effect" and "indirect effect"; we also introduce a new causal estimand that indicates a non-zero "infectiousness effect" in the context of vaccine trials (Datta et al., 1999). Finally, in section 7, we offer some concluding remarks and directions for future research.

2 Preliminaries

2.1 Counterfactuals

As in Hudgens and Halloran (2008), suppose data is observed on $N > 1$ groups of individuals, or blocks of units. For $i = 1, \ldots, N$ let n_i denote the number of individuals in group i and let $\mathbf{A}_i \equiv (A_{i1}, \ldots, A_{in_i})$ denote the treatments those n_i individuals received. Throughout, we assume perfect compliance, that is treatment assigned to an individual is equivalent to treatment received by the individual. We assume that A_{ij} is a dichotomous random variable with support equal to $\{0, 1\}$, so that $\mathbf{A}_i$ takes values in the set $\{0, 1\}^{n_i}$. Let $\mathbf{A}_{i,-j} \equiv (A_{i1}, \ldots A_{in_i}) \backslash A_{ij} \equiv (A_{i1}, \ldots, A_{ij-1}, A_{ij+1}, \ldots A_{in_i})$ denote the $n_i - 1$ subvector of $\mathbf{A}_i$ with the *jth* entry deleted. Following Hudgens and Halloran (2008) and Sobel (2006), we refer to $\mathbf{A}_i$ as an intervention, treatment or allocation program, to distinguish it from the individual treatment A_{ij} Furthermore, for $n = 1, 2, \ldots$, we define $\mathscr{A}(n)$ as the set of vectors of possible treatment allocations of length n; for instance $\mathscr{A}(2) \equiv \{(0, 0), (0, 1), (1, 0), (1, 1)\}$. Therefore, $\mathbf{A}_i$ takes one of 2^{n_i} possible values in $\mathscr{A}(n_i)$, while $\mathbf{A}_{i,-j}$ takes values in $\mathscr{A}(n_i - 1)$ for all j. For positive integers n and k, we further define $\mathscr{A}(n, k)$ to be the subset of $\mathscr{A}(n)$ wherein exactly k individuals receive treatment 1, that is every element $\mathbf{a}$ of $\mathscr{A}(n, k)$ satisfies $1_n^T \mathbf{a} = k$, where $\mathbf{1}_n$ is the vector of length n with entries all equal to one.

For each block i, we shall assume there exist counterfactual (potential outcome) data $\mathbf{Y}_i(\cdot) = \{\mathbf{Y}_i(\mathbf{a}_i) : \mathbf{a}_i \in \mathscr{A}\}$ where $\mathbf{Y}_i(\mathbf{a}_i) = \{Y_{i1}(\mathbf{a}_i), \ldots, Y_{in_i}(\mathbf{a}_i)\}$, and $Y_{ij}(\mathbf{a}_i)$ is individual j's response under treatment allocation $\mathbf{a}_i$; and that the observed outcome Y_{ij} for individual j in block i is equal to his counterfactual outcome $Y_{ij}(\mathbf{A}_i)$ under the realized treatment allocation $\mathbf{A}_i$. The notation $Y_{ij}(\mathbf{a}_i)$ makes explicit the possibility for interference between individuals within a block, that is, the potential outcome for individual j may depend on another's individual treatment assignment in block j. Also, note that for counterfactuals to remain well defined, this notation implicitly assumes that counterfactuals for an individual in block i do not depend on treatment assignments of individuals in a different block $i' \neq i$. This encodes the assumption of partial interference considered by Sobel (2006) and Hudgens and Halloran (2008), which they point out to be particularly appropriate when the observed blocks are well separated by space or time such as in some group randomized studies in the social sciences, or in some community-randomized vaccine trials. The ordinary no interference assumption (Cox, 1958; Rubin, 1980) generally made in the causal inference literature is then that for all i and j if $\mathbf{a}_i$ and $\mathbf{a}_i'$ are such that $a_{ij} = a_{ij}'$ then $Y_{ij}(\mathbf{a}_i) = Y_{ij}(\mathbf{a}_i')$, which in turn implies that the counterfactual outcomes for individual j in group i can be written as $\{Y_{ij}(a) : a = 0, 1\}$.

Hereafter, we follow the convention in Sobel (2006) and Hudgens and Halloran (2008), and suppose that $\mathbf{Y}_i(\cdot)$ is fixed as it does not depend on the random treatment allocation program $\mathbf{A}_i$. In addition to treatment and outcome data, we suppose that we also observe fixed data $\mathbf{L}_i$

$= (L_{il}, \ldots, L_{in_i})$, $i = 1, \ldots N$, where L_{ij} denotes pretreatment covariates for individual i in block j; we allow L_{ij} to contain block level covariate along with block aggregates of individual level covariates.

2.2 Treatment Assignment in Group Randomized Experiments

In group randomized experiments, treatment allocation is determined by the experimenter; therefore the assignment mechanism $\pi_i(\mathbf{A}_i)$ of $\mathbf{A}_i$ is known. Let $\pi_i(\mathbf{A}_i; \alpha_0)$ denote an experimenter's particular choice of parametrization for the distribution of $\mathbf{A}_i$ indexed by the parameter α_0, that is $\pi_i(\mathbf{A}_i) = \pi_i(\mathbf{A}_i; \alpha_0)$. In this paper, we consider two types of parametrizations.

Definition—*(A) A parametrization of type A with parameter n_i and $K_{0,i}$ for block i, entails a so-called mixed individual group assignment strategy, whereby the treatment program $\mathbf{A}_i$ in block i is randomly allocated conditional on $1_n^T \mathbf{A}_i = \sum_{j=1}^{n_i} A_{ij} = K_{0,i}$ with probability mass function*

$$\pi_i(\mathbf{A}_i;\alpha_0) = I(A_i \in A(n_i, K_{0,i}))/\begin{pmatrix} n_i \\ K_{0,i} \end{pmatrix}$$

(B) A parametrization of type B entails a Bernoulli individual group assignment strategy, whereby treatment is randomly assigned to different individuals within block i according to the known probability mass function

$$\pi_i(\mathbf{A}_i;\alpha_0) = \prod_{j=1}^{n_i} \alpha_0^{A_{ij}} (1-\alpha_0)^{1-A_{ij}}$$

where $0 < \alpha_0 < 1$.

For example, two type A treatment assignment strategies α_0 and α_1 might entail randomly assigning half of n_i individuals in group i to treatment 1 and the other half to treatment 0 under a strategy corresponding to α_0 versus assigning all individuals in a group to treatment zero under the second strategy corresponding to α_1. Similarly, two treatment assignment strategies $\alpha_0^* \text{and} \alpha_1^*$ of the second type might assign each individual in a group to treatment 1 with probability 1/2 under strategy α_0^* versus assigning each individual in a group to treatment 0 with probability 1/3 under strategy α_1^*. Sobel (2006) and Hudgens and Halloran (2008) considered Type A treatment allocation programs in group randomized trials; in Section 5, we show that allocation programs of type (B) play an important conceptual role in the Definition and estimation of causal effects in observational studies.

Suppose our goal is to assess the causal effects of assigning groups to α_0, compared to α_1, where α_0 and α_1 are two individual group assignment strategies of type A. To achieve this goal in an experimental study, Hudgens and Halloran (2008) considered the following two-stage group randomization framework. In the first stage, each of the N groups is randomly assigned to either α_0 or α_1. In the second stage individuals within a group are randomly

assigned to treatment conditional on their group's assignment in the first stage. For instance, in the first stage, half of the *N* groups might be assigned to an allocation strategy α_0 while the other half is assigned to α_1; in the second stage, two-thirds of the individuals within groups assigned α_0 are randomly assigned to treatment 1, while one-third of the individuals within a group assigned to α_1 receive treatment 1. Such a design is commonly known as split-plot randomization or pseudo-cluster randomization. As Hudgens and Halloran (2008) point out, two-stage randomization designs are key to obtaining answers for important public health questions in the face of interference, such as: how many cases due to an infectious disease will be averted by vaccinating two-thirds of the population compared to only vaccinating one-third of the population?

3 Causal Estimands

3.1 Direct Causal Effects

Following Halloran and Struchiner (1995), we define the individual direct causal effect of treatment 0 compared to treatment 1 for individual *j* in group *i* by:

$$DE_{ij}(\mathbf{a}_{i,-j}) \equiv Y_{ij}(\mathbf{a}_{i,-j}, a_{ij}{=}0) - Y_{ij}(\mathbf{a}_{i,-j}, a_{ij}{=}1)$$

and the individual average direct causal effect for individual *j* in group *i* by

$$\overline{DE}_{ij}(\alpha_0) \equiv \overline{Y}_{ij}(0;\alpha_0) - \overline{Y}_{ij}(1;\alpha_0) \quad (1)$$

where for $a = 0, 1$,

$$\overline{Y}_{ij}(a;\alpha_0) \equiv \sum_{s\in\mathscr{A}(n-1)} Y_{ij}(\mathbf{a}_{i,-j}{=}\mathbf{s}, a_{ij}{=}a)\Pr_{\alpha_0}(\mathbf{A}_{i,-j}{=}\mathbf{s}|A_{ij}{=}\alpha)$$

$$\text{with } \Pr_{\alpha_0}(\mathbf{A}_{i,-j}{=}\mathbf{s}|A_{ij}{=}\alpha){=}\frac{\pi_i(\mathbf{A}_{i,-j}{=}\mathbf{s},\alpha;\alpha_0)}{\sum_{\mathbf{s}'\in\mathscr{A}(n-1)}\pi_i(\mathbf{s}',a;\alpha_0)}$$

Note that in the above display, and until stated otherwise, $\pi_i(\cdot;\alpha_0)$ may either be of Type A or B. Thus, $\overline{DE}_{ij}(\alpha_0)$ is a difference in individual average counterfactual outcomes when $a_{ij} = 0$ and when $a_{ij} = 1$ under α_0. This is a marginal causal effect as it is a comparison between expected values of the marginal distributions of $Y_{ij}(\mathbf{A}_{i,-j}, a_{ij} = 0)$ and of $Y_{ij}(\mathbf{A}_{i,-j}, a_{ij} = 1)$ with respect to α_0. Finally, we define the group average direct causal effect by $\overline{DE}_i(\alpha_0){=}\sum_{j=1}^{n_i}\overline{DE}_{ij}(\alpha_0)/n_i$ and the population average direct causal effect by $\overline{DE}(\alpha_0){=}\sum_{i=1}^{N}\overline{DE}_i(\alpha_0)/N$.

3.2 Indirect Causal Effects or"Spillover Effects"

Halloran and Struchiner (1995) also define an individual indirect causal effect as the causal effect on an individual of the treatment received by others in the group. Specifically, let

$IE_{ij}(\mathbf{a}_{i,-j}, \mathbf{a}'_{i,-j})$ be the individual indirect causal effect on subject *j* in group *i* of treatment allocation $\mathbf{a}_i$ compared with $\mathbf{a}'_i$ so that:

$$IE_{ij}(\mathbf{a}_{i,-j}, \mathbf{a}'_{i,-j}) = Y_{ij}(\mathbf{a}_{i,-j}, \alpha_{ij}=0) - Y_{ij}(\mathbf{a}_{i,-j}, \alpha'_{ij}=0)$$

Sobel (2006) refers to the indirect effect defined above as a"spillover effect." Note that if interference is absent then $IE_{ij}(\mathbf{a}_{i,-j}, \mathbf{a}'_{i,-j})=0$. Similar to direct effects, define the individual average indirect causal effect by $\overline{IE}_{ij}(\alpha_0, \alpha_1)=\overline{Y}_{ij}(0, \alpha_0) - \overline{Y}_{ij}(1;\alpha_1)$. Finally, define the group average indirect causal effect as $\overline{IE}_i(\alpha_0, \alpha_1)=\sum_{j=1}^{n_i} \overline{IE}_{ij}(\alpha_0, \alpha_1)/n_i$ and the population average indirect causal effect as $\overline{IE}(\alpha_0, \alpha_1)=\sum_{i=1}^{N} \overline{IE}_i(\alpha_0, \alpha_1)/N$.

3.3 Total Causal Effects

Total effects reflect both the direct and the indirect effects of a particular treatment assignment on an individual. Following Halloran and Struchiner (1995) we define the individual total causal effects for individual *j* in group *i* as:

$$TE_{ij}(\mathbf{a}_{i,-j}, \mathbf{a}'_{i,-j}) \equiv Y_{ij}(\mathbf{a}_{i,-j}, \alpha_{ij}=0) - Y_{ij}(\mathbf{a}'_{i,-j}, \alpha_{ij}=1),$$

the individual average total causal effect by $\overline{TE}_{ij}(\alpha_0, \alpha_1)=\overline{Y}_{ij}(0, \alpha_0) - \overline{Y}_{ij}(1;\alpha_1)$, the group average total causal effect by $\overline{TE}_i(\alpha_0, \alpha_1)=\sum_{j=1}^{n_i} \overline{TE}_{ij}(\alpha_0, \alpha_1)/n_i$ and the population average total causal effect by $\overline{TE}(\alpha_0, \alpha_1)=\sum_{i=1}^{N} \overline{TE}_i(\alpha_0)/N$.

3.4 Overall Causal Effects

Following Hudgens and Halloran (2008), we define the individual overall causal effect of treatment $\mathbf{a}_i$ compared to treatment $\mathbf{a}'_i$ for individual *j* in group *i* by

$$\overline{OE}_{ij}(\mathbf{a}_i, \mathbf{a}'_i)=Y_{ij}(\mathbf{a}_i) - Y_{ij}(\mathbf{a}'_i)$$

Similarly, define the individual average overall causal effect comparing α_0 to α_1 by $\overline{OE}_{ij}(\alpha_0, \alpha_1)=\overline{Y}_{ij}(\alpha_0) - \overline{Y}_{ij}(\alpha_1)$, the group average overall causal effect by $\overline{OE}_i(\alpha_0, \alpha_1)=\sum_{j=1}^{n_i} \overline{OE}_{ij}(\alpha_0, \alpha_1)/n_i$ and the population average overall effect by $\overline{OE}(\alpha_0, \alpha_1)=\sum_{i=1}^{N} \overline{OE}_i(\alpha_0)/N$

The following simple yet instructive properties describe the relationship between the various causal effects:

1. It follows immediately from their Definitions, that total effects at the individual, group or population levels can be decomposed as the sum of direct and indirect

causal effects at the corresponding level. That is, for example $\overline{TE}(\alpha_0,\alpha_1)=\overline{DE}(\alpha_1)+\overline{IE}(\alpha_0,\alpha_1)$ (Hudgens and Halloran, 2008).

2. Total causal effects are not commutative, for instance $\overline{TE}(\alpha_0,\alpha_1) \neq \overline{TE}(\alpha_1,\alpha_0)$. However $\overline{IE}(\alpha_0,\alpha_1)=-\overline{IE}(\alpha_1,\alpha_0) \Rightarrow \overline{DE}(\alpha_0)+\overline{DE}(\alpha_1)=\overline{TE}_{ij}(\alpha_0,\alpha_1)+\overline{TE}_{ij}(\alpha_1,\alpha_0)$, so that while the total causal effects are not necessarily equal, they are constrained in sum to equal the sum of direct effects (Hudgens and Halloran, 2008).

3. If $\overline{IE}(\alpha_0,\alpha_1)=\overline{IE}(\alpha_1,\alpha_0)=0$, then $\overline{TE}(\alpha_0,\alpha_1)=\overline{TE}(\alpha_1,\alpha_0) \iff \overline{DE}(\alpha_0)=\overline{DE}(\alpha_1)$. In the absence of indirect effects, the total effects are commutative if and only if the direct effects are equal (Hudgens and Halloran, 2008).

   We also have the following decomposition for the overall effect:

4. The group average overall effects are equal to a weighted sum of the group average indirect, direct and total effects: $\overline{OE}_i(\alpha_0,\alpha_1)=\Pr(A_{ij}=0;\alpha_1)\overline{IE}_i(\alpha_0,\alpha_1)+\Pr(A_{ij}=1;\alpha_1)\overline{TE}_i(\alpha_0,\alpha_1)+\Pr(A_{ij}=1;\alpha_0)\overline{DE}_i(\alpha_1,\alpha_0)$, where

$$\Pr(A_{ij}=a_{ij};\alpha)=\sum_{\mathbf{s}\in\mathscr{A}(n-1)}\pi_i(\mathbf{s},a_{ij};\alpha)=\begin{cases}\frac{K_i}{n_i} & \text{if } \pi_i(\mathbf{A}_i;\alpha)\text{is of type A}\\ \alpha & \text{if } \pi_i(\mathbf{A}_i;\alpha)\text{is of type B}\end{cases}$$

Under the assumption of no interference between individuals of a group, the individual indirect causal effect is equal to zero and therefore individual, group and population average causal total effects are equal to the average causal direct effects at the corresponding level. Recall that in the absence of interference, the counterfactual outcomes for individual *j* in group *i* can be written as $\{Y_{ij}(a) : a = 0, 1\}$ and the individual and group average causal effect ct become $Y_{ij}(1) - Y_{ij}(0)$ and $\sum_{j=1}^{n_i}\{Y_{ij}(1)-Y_{ij}(0)\}/n_i$ respectively. Furthermore, the assumption of no interference implies that the various causal effects do not depend on the treatment assignment strategies α_0 and α_1, whereas in the presence of interference within groups, these effects do in general depend on the assignment strategies.

4 Inference in group randomized studies

4.1 Estimation

In this section, we consider the estimation of the following four key causal contrasts, the population average direct causal effect $\overline{DE}(\alpha_0)$, the population average indirect causal effect $\overline{IE}(\alpha_0,\alpha_1)$, the population average total causal effect $\overline{TE}(\alpha_0,\alpha_1)$ and the population average overall effect $\overline{OE}(\alpha_0,\alpha_1)$. Unbiased estimators of these parameters under a two-stage randomization scheme were proposed by Hudgens and Halloran (2008) under the following assumption:

Assumption 1—*Let* $\mathbf{S} \equiv (S_1, \ldots, S_N)$ *denote the first stage of randomization group assignments with* $S_i = 1$ *if group i is assigned to* α_0 *and zero if group i is assigned to* α_1. *Let* η *denote the parametrization for the distribution of S and let* $C = \sum_i S_i$ *denote the number of groups assigned* α_1. *Then,* $\{\eta, \alpha_0, \alpha_1\}$ *are assumed to be Type A parametrizations.*

Suppose $S_i = 1$ and let $\hat{Y}_i(\alpha;\alpha_0)=\sum_{j=1}^{n_i} I(A_{ij}=a)Y_{ij}(\mathbf{A}_i)/\sum_{j=1}^{n_i} I(A_{ij}=a)$, also define $\hat{Y}_i(\alpha;\alpha_0)=\sum_{i=1}^{N} \hat{Y}_i(a;\alpha_0)I(S_i=1)/\sum_{i=1}^{N} I(S_i=1)\hat{Y}_i(\alpha_0)=\sum_{j=1}^{n_i} Y_{ij}(\mathbf{A}_i)/n_i$, and $\hat{Y}(\alpha_0)=\sum_{i=1}^{n_i} \hat{Y}_i(\alpha_0)I(S_i=1)/\sum_{i=1}^{N} I(S_i=1)$. Hudgens and Halloran (2008) proposed the following estimators:

$$\hat{DE}(\alpha_0)=\hat{Y}(0;\alpha_0) - \hat{Y}(1;\alpha_0), \quad (2)$$

$$\hat{IE}(\alpha_0,\alpha_1)=\hat{Y}(0;\alpha_0) - \hat{Y}(0;\alpha_1), \quad (3)$$

$$\hat{TE}(\alpha_0,\alpha_1)=\hat{Y}(1;\alpha_0) - \hat{Y}(0;\alpha_1), \quad (4)$$

$$\hat{OE}(\alpha_0,\alpha_1)=\hat{Y}(\alpha_0) - \hat{Y}(\alpha_1), \quad (5)$$

which they showed to be unbiased under Assumption 1, i.e.

$$E\{\hat{DE}(\alpha_0)\}=\overline{DE}(\alpha_0),$$

$$E\{\hat{IE}(\alpha_0,\alpha_1)\}=\overline{IE}(\alpha_0,\alpha_1),$$

$$E\{\hat{TE}(\alpha_0,\alpha_1)\}=\overline{TE}(\alpha_0,\alpha_1),$$

$$E\{\hat{OE}(\alpha_0,\alpha_1)\}=\overline{OE}(\alpha_0,\alpha_1)$$

where the expectation is taken with respect to the joint density of $(\mathbf{S},\mathbf{A}_1, \ldots, \mathbf{A}_N)$.

4.2 Variance Estimation

4.2.1 Variance Estimation under Stratified interference—Unbiased estimation of the variances of the various estimators of the previous section appears not to be generally available without additional assumptions regarding the underlying structure of interference. Hudgens and Halloran (2008) illustrate this difficulty by considering the estimation of *Var* $(\hat{Y}(1;\alpha_0)|S_i=\mathbf{1})$ under assumption 1 only. They note that the estimator $\hat{Y}(1;\alpha_0)$ is based on a single systematic random sample of fixed size K_i from the set of potential outcomes $\{Y_{ij}$

$(\mathbf{a}_i) : \mathbf{a}_i \in \mathscr{A}(n_i;K_i), z_{ij} = 1\}$. By the non-existence of an unbiased estimator of the variance of the sample mean from a single systematic sample, this implies the non-existence of an unbiased estimator of $Var(\hat{Y}(1;\alpha_0)|S_i=\mathbf{1})$. However, as we show in the next lemma, the non-existence of an unbiased estimator of $Var(\hat{Y}(1;\alpha_0)|S_i=\mathbf{1})$ does not preclude the possibility for simple yet conservative estimation of the latter quantity, as an unbiased estimator of an upper bound for the variance is often a useful measure of uncertainty. The following lemma gives the result for a nonnegative outcome.

Lemma 1: *Suppose that* $Y_{ij}(\mathbf{a}_i) \geq 0$ *for all* $\mathbf{a}_i \in A(n_i;K_{0,i})$ and for $j = 1, \ldots, n_i$, *and define* $\hat{Var}_u(\hat{Y}_i(1;\alpha_0)|S_i=1) \equiv$

$$\frac{1-\pi_i(\mathbf{A}_i;\alpha_0)}{K_{0,i}^2}\left\{\sum_{j=1}^{n_i}A_{ij}Y_{ij}^2(\mathbf{A}_i)+\sum_{j\neq j'}^{n_i}A_{ij}A_{ij'}Y_{ij}(\mathbf{A}_i)Y_{ij'}(\mathbf{A}_i)\right\},$$

then the following holds under Assumption 1:

$$E\{\hat{Var}_u(\hat{Y}_i(1;\alpha_0)\,\big|S_i=1)\big|S_i=1\} \geq Var(\hat{Y}_i(1;\alpha_0)\,\big|\,S_i=1)$$

The proof of this lemma is given in the appendix.

In contrast with Lemma 1 Hudgens and Halloran (2008) consider variance estimators that rely on the following assumption of Stratified interference.

Assumption 2: *Stratified interference: For*

$k=1,\ldots,n_i-1, Y_{ij}(\mathbf{a}_i)=Y_{ij}(\mathbf{a}_i^{'})$ *for all* $\mathbf{a}_i, \mathbf{a}_i^{'} \in \mathscr{A}(n_i,k)$, *such that* $a_{ij}=a_{ij}^{'}$

Assumption 2 states that $\mathbf{a}_i \mapsto Y_{ij}(\mathbf{a}_i)$ is a function of $\mathbf{a}_i$ only through $(a_{ij}, \sum j' \neq j\ a_{ij'})$, that is an individual's counterfactual outcome only depends on his exposure level a_{ij}, and on the total number of people exposed in his group. Let $Y_{ij}(a_{ij};\alpha_0) \equiv Y_{ij}(a_{ij}, \mathbf{a}_{i,-j};\alpha_0)$ for any $\mathbf{a}_{i,-j} \in A(n_i - 1, K_i - a_{ij})$, $a_{ij} = 0, 1$; and let

$$\hat{\sigma}_{ia}^2(\alpha) \equiv \sum_{j=1}^{n_i}[Y_{ij}(a;\alpha) - \hat{Y}_i(a;\alpha)]^2 1(A_{ij}=a)/(K_i - 1)$$

$$\hat{\sigma}_{ga}^2(\alpha) \equiv \sum_{i=1}^{N}[\hat{Y}_i(a;\alpha) - \hat{Y}(a;\alpha)]^2 S_i/(C - 1)$$

where $\hat{\sigma}_{ia}^2(\alpha)$ is the within-group sample variance and $\hat{\sigma}_{ga}^2(\alpha)$ the between group sample variance for individuals with $A_{ij} = a \in \{0, 1\}$. Also, let

$$\hat{\sigma}^2_{DE}(\alpha) \equiv \sum_{i=1}^{N} [\hat{DE}_i(\alpha) - \hat{DE}(\alpha)]^2 S_i/(C-1)$$

$$\hat{\sigma}^2_{M}(\alpha_0) \equiv \sum_{i=1}^{N} [Y_i(\alpha) - \hat{Y}(\alpha)]^2 S_i/(C-1)$$

and

$$\hat{Var}\{\hat{DE}_i(\alpha)|S_i=1\} = \frac{\hat{\sigma}^2_{i1}(\alpha)}{K_i} + \frac{\hat{\sigma}^2_{i0}(\alpha)}{n_i - K_i}$$

and define

$$\hat{Var}\{\hat{DE}(\alpha_0)\Big|S_i=1\} \equiv (1-\frac{C}{N})\hat{\sigma}^2_{DE}(\alpha_0) + \frac{1}{CN}\sum_{i=1}^{N}\hat{Var}\{\hat{DE}_i(\alpha_0)\Big|S_i=1\}S_i \quad (6)$$

$$\hat{Var}\{\hat{IE}(\alpha_0,\alpha_1)\} \equiv \frac{\hat{\sigma}^2_{g0}(\alpha_0)}{N-C} + \frac{\hat{\sigma}^2_{g0}(\alpha_1)}{C} \quad (7)$$

$$\hat{Var}\{\hat{TE}(\alpha_0,\alpha_1)\} \equiv \frac{\hat{\sigma}^2_{g0}(\alpha_0)}{N-C} + \frac{\hat{\sigma}^2_{g1}(\alpha_1)}{C} \quad (8)$$

$$\hat{Var}\{\hat{OE}(\alpha_0,\alpha_1)\} \equiv \frac{\hat{\sigma}^2_{M}(\alpha_0)}{N-C} + \frac{\hat{\sigma}^2_{M}(\alpha_1)}{C} \quad (9)$$

Hudgens and Halloran (2008) proved that under assumptions 1 and 2:

$$E[\hat{Var}\{\hat{DE}(\alpha_0)\Big|S_i=1\}] \geq Var\{\hat{DE}(\alpha_0)\Big|S_i=1\} \quad (10)$$

$$E[\hat{Var}\{\hat{IE}(\alpha_0,\alpha_1)\}] \geq Var\{\hat{IE}(\alpha_0,\alpha_1)\} \quad (11)$$

$$E[\hat{Var}\{\hat{TE}(\alpha_0,\alpha_1)\}] \geq Var\{\hat{TE}(\alpha_0,\alpha_1)\} \quad (12)$$

$$E[\hat{Var}\{\hat{OE}(\alpha_0,\alpha_1)\}] \geq Var\{\hat{OE}(\alpha_0,\alpha_1)\} \quad (13)$$

That is the variance estimators (6)-(9) are generally conservative. However, as they show in equation (10), equality holds if and only if

$$Y_{ij}(1;\alpha_0)=Y_{ij}(0;\alpha_0)+\psi_{D,i} \quad (14)$$

for fixed constant, for $j = 1, \ldots, n_i$ and $i = 1, \ldots, N$, which is equivalent to an additive individual direct causal effect across all groups. Note that when Y_{ij} ($\mathbf{A}_i$) is binary, and $0 < |DE(\alpha_0)| < 1$, then the hypothesis of additive direct treatment effects cannot hold as the only values of $DE(\alpha_0)$ consistent with additivity are 0, 1 and −1. Hudgens and Halloran (2008) also establish analogous conditions under which equality holds for each of the other equations (11)–(13).

Despite the availability under assumptions 1 and 2, of reasonable variance estimators given by equations (6) – (9) for the various estimators of causal effects proposed by Hudgens and Halloran (2008), a formal framework for statistical inference on population average causal effects is currently lacking. As a remedy, in the following section, we develop a finite sample framework for making causal inferences in the context of interference.

4.3 Finite sample inference for a binary outcome

We construct novel finite sample confidence intervals for the four population average causal effects of interest. To simplify the exposition, we mainly focus on the case of a binary outcome. To the best of our knowledge there currently exists no method, whether finite or large sample-based, to construct a confidence interval for any of the causal parameters of current interest. In a technical report, we show that $\hat{DE}(\alpha_0) - DE(\alpha_0)$ admits an alternative representation as a martingale, an observation which enables us to use a Hoeffding-type exponential inequality to obtain the desired finite sample confidence interval. We prove the following results.

Theorem 1—*For any level* $\gamma \in (0, 1)$, *the interval*

$$C_{DE}(\gamma, p(\alpha_0), q, N) \equiv (\hat{DE}(\alpha_0) - \varepsilon^*_{DE}(q,\gamma,N), \hat{DE}(\alpha_0) + \varepsilon^*_{DE}(q,\gamma,N))$$

is a finite sample $(1-\gamma)$ *CI of DE* (α_0) *under assumption 1, where*

$$\varepsilon^*_{DE}(q,\gamma,N) = \sqrt{\frac{\left[4(\frac{1}{q}-1)^2 + \frac{\sum_{i=1}^{N}\left(\frac{L_{DE,i}}{q}\right)^2}{N}\right]}{2N}\ln(\frac{2}{\gamma})} \quad (15)$$

$q \equiv \Pr(S_i=1) = \frac{C}{N}$ and *for* $i = 1, \ldots, N$

$$L_{DE,i}(\alpha_0) \equiv 2\left(1 - \frac{1}{\binom{n_i}{K_{0,i}}}\right).$$

According to the theorem, for each value of (q, N, γ), the coverage probability $\Pr\{DE(\alpha_0) \in C_{DE}(\gamma, q, N)\}$ is guaranteed under assumption 1 to be no smaller than 95%, with the length of $C_{DE}(\gamma, q, N)$ proportional to $\frac{1}{N^{1/2}}$, so that for a fixed value of (γ q), $C_{DE}(\gamma\, q, N)$ becomes increasingly precise as the number of groups in the study grows. However, we note that $C_{DE}(\gamma\, q, N)$ may not be particularly useful when N is small, for those values of (γ q) such that $\varepsilon^*_{DE}(q, \gamma, N) \geq 2$. This is because in such a case, the corresponding confidence interval is noninformative, as it contains the entire range of possible values of $\overline{DE}(\alpha_0)$, since $[-1, 1] \subseteq C_{DE}(\gamma, q, N)$ and $|\overline{DE}(\alpha_0)| \leq 1$. To further illustrate this point, suppose that $\frac{1}{\binom{n_i}{K_{0,i}}} \approx 0$ and $q = 1/2$, then $\varepsilon^*_{DE}(1/2, \gamma, N) \approx \frac{6}{\sqrt{N}}$. This implies that $C_{DE}(\gamma\, q, N)$ is guaranteed to be noninformative for values of $N \leq 9$. As made evident in the proof of the theorem, the term $4\left(\frac{1}{q} - 1\right)^2$ in equation (15) is an upper bound for the squared absolute deviation of the conditional average direct effect $E\{\hat{\mathrm{DE}}(\alpha_0)|S{=}1\} = \frac{1}{c}\sum_{i:S_i=1} \overline{Y}_i(0;\alpha_0) - \overline{Y}_i(1;\alpha_0)$ from the population average direct effect $\overline{Y}(0; \alpha_0) - \overline{Y}(1; \alpha_0)$. This bound increases as q decreases towards zero, a situation which can arise in a study where the proportion of groups randomized to the treatment allocation α_0 is very small, and can happen even when C and N are both relatively large. This will invariably result in an increase in uncertainty in our inferences on $\overline{DE}(\alpha_0)$. However, we note that more accurate inferences may still be possible for the population conditional average causal direct effect which we define as

$$\overline{DE}_c(\alpha_0) \equiv \frac{1}{C}\sum_{i:S_i=1} \overline{Y}_i(0;\alpha_0) - \overline{Y}_i(1;\alpha_0)$$

and which corresponds to the average causal direct effect for the population of groups actually randomized to α_0. The next theorem provides a finite sample confidence interval for $\overline{DE}_c(\alpha_0)$.

Theorem 2—*For any level* $\gamma \in (0,1)$, *the interval*

$$C_{DE_c}(\gamma, q, N) \equiv (\hat{DE}(\alpha_0) - \varepsilon^*_{DE_c}(q, \gamma, N), \hat{DE}(\alpha_0) - \varepsilon^*_{DE_c}(q, \gamma, N)$$

is a finite sample $(1 - \gamma)$ *CI of* $\overline{DE}(\alpha_0)$ *under assumption 1, where*

$$\varepsilon^*_{DE_c}(q, \gamma, N) = \frac{1}{\sqrt{C}}\sqrt{\left[\sum_{i:S_i=1} (L_{DE,i}(q))^2/2C\right]\ln\left(\frac{2}{\gamma}\right)}$$

Note that both $C_{DE}(\gamma, q, N)$ and $CI_{DE_c}(\gamma, q, N)$ are centered around the same estimator $\hat{DE}(\alpha_0)$, which is unbiased for $\overline{DE}(\alpha_0)$ and is conditionally unbiased for $\overline{DE}_c(\alpha_0)$. However, the length of the second confidence interval no longer includes the term $4(\frac{1}{q}-1)^2$ and thus will often be substantially shorter.

The following theorem provides a finite sample confidence interval for the population average indirect causal effect.

Theorem 3—*For any level* $\gamma \in (0, 1)$, *the interval*

$$C_{IE}(\gamma) \equiv (\hat{IE}(\alpha_0, \alpha_1) - \varepsilon^*_{IE}(q, \gamma, N), \hat{IE}(\alpha_0, \alpha_1) + \varepsilon^*_{IE}(q, \gamma, N)$$

is a finite sample $(1-\gamma)CI$ *of* $\overline{IE}(\alpha_0, \alpha_1)$ under assumption 1, where

$$\varepsilon^*_{IE}(q, \gamma, N) = \sqrt{\frac{\left[\max\left\{\frac{1}{q^2}, \frac{1}{(1-q)^2}\right\} + \sum_i L^2_{IE,i,}(q)/N\right]}{2N} \ln\left(\frac{2}{\gamma}\right)}$$

and

$$L_{IE,i}(q) = \max\left\{\frac{1}{q}\left(1 - \frac{1}{\binom{n_i}{K_{0,i}}}\right), \frac{1}{1-q}\left(1 - \frac{1}{\binom{n_i}{K_{1,i}}}\right)\right\}$$

The next two theorems give finite sample confidence intervals for the population average total causal effect and for the population average overall causal effect respectively.

Theorem 4—*For any level* $\gamma \in (0, 1)$, *the interval*

$$C_{TE}(\gamma) \equiv (\hat{TE}(\alpha_0, \alpha_1) - \varepsilon^*_{TE}(q, \gamma, N), \hat{TE}(\alpha_0, \alpha_1) + \varepsilon^*_{TE}(q, \gamma, N)$$

is a finite sample $(1-\gamma)CI$ *of* $\overline{TE}(\alpha_0, \alpha_1)$ *under assumption 1, where*

$$\varepsilon^*_{TE}(q, \gamma, N) = \sqrt{\frac{\left[\max\left\{\frac{1}{q^2}, \frac{1}{(1-q)^2}\right\} + \sum_i L^2_i(q)/N\right]}{2N} \ln\left(\frac{2}{\gamma}\right)}$$

and

$$L_{TE,i}(q)=\max\left\{\frac{1}{q}\left(1-\frac{1}{\binom{n_i}{K_{0,i}}}\right),\frac{1}{1-q}\left(1-\frac{1}{\binom{n_i}{K_{1,i}}}\right)\right\}$$

Theorem 5—*For any level* $\gamma \in (0, 1)$*, the interval*

$$C_{OE}(\gamma,q,N)\equiv(\hat{OE}(\alpha_0,\alpha_1)-\varepsilon^*_{OE}(q,\gamma,N),\hat{OE}(\alpha_0,\alpha_1)+\varepsilon^*_{OE}(q,\gamma,N)$$

is a finite sample $(1-\gamma)CI$ *of* $\overline{OE}(\alpha_0,\alpha_1)$ *under assumption 1, where*

$$\varepsilon^*_{OE}(q,\gamma,N)=\sqrt{\frac{\left[\max\left\{\frac{1}{q^2},\frac{1}{(1-q)^2}\right\}+\sum_i L_i^2(q)/N\right]}{2N}\ln\left(\frac{2}{\gamma}\right)}$$

and

$$L_{OE,i}(q)=\max\left\{\frac{1}{q}\left(1-\frac{1}{\binom{n_i}{K_{0,i}}}\right),\frac{1}{1-q}\left(1-\frac{1}{\binom{n_i}{K_{1,i}}}\right)\right\}$$

Note that $\varepsilon^*_{IE}(q,\gamma,N)=\varepsilon^*_{TE}(q,\gamma,N)=\varepsilon^*_{OE}(q,\gamma,N)$, with the corresponding confidence intervals having identical length. Future work could improve about the length of these confidence intervals by a sharpening of the exponential inequalities used in their derivation (van der Vaart and Wellner, 1996) and by leveraging additional assumptions such as that of Stratified interference or by deriving potentially sharper alternative exponential inequalities. In future work, we also plan to consider inference for continuous and possibly unbounded outcomes. The technical developments necessary to achieve these results are beyond the scope of the current paper and will be addressed elsewhere.

5 Towards Inference in observational studies

In this section, we briefly consider an approach for drawing causal inferences from observational data in the presence of interference. We begin by noting that in the absence of (two-stage) randomization, the estimators of Section 5 are no longer valid in an observational study. This is because Assumption 1 is in general no longer tenable in the non-experimental setting of an observational study, therefore, a different approach is needed. To make progress, we consider the following assumption:

Assumption 3

For $i = 1, \ldots, N$*, we assume that conditional on* L_i*, the treatment allocation* A_i *is independent of the counterfactual variables* $Y_i(\cdot)$*, that is:*

$$\Pr\{\mathbf{A}_i=\mathbf{a}_i \left| \mathbf{L}_i, \boldsymbol{Y}_i(\cdot)\right\}=f_{\mathbf{A}\left|\mathbf{L},i\right.}(\mathbf{a}_i \left| \mathbf{L}_i\right) \quad (16)$$

where $f_{\mathbf{A}|\mathbf{L}i}(\mathbf{a}_i|\mathbf{L}_i) \equiv \Pr\{\mathbf{A}_i = \mathbf{a}_i|\mathbf{L}_i\}$

This assumption is a group-level generalization of the standard conditional randomization assumption routinely made at the individual-level in the analysis of observational studies. It states that the treatment allocation program $\mathbf{A}_i$ is randomly assigned to individuals in group *i* conditional on the vector of covariates $\mathbf{L}_i$ observed on these individuals. Whereas in the previous section, the outcome was assumed to be binary, hereafter, no such assumption is needed. In addition to Assumption 3, we suppose that the following positivity assumption holds:

Assumption 4

For $i = 1, \ldots, N$ *we assume that conditional on* $\mathbf{L}_i$, *we have that for all* $\mathbf{a}_i \in \mathscr{A}(n_i)$

$$\Pr\{\mathbf{A}_i=\mathbf{a}_i\left|\mathbf{L}_i\right\}>0 \quad (17)$$

Assumption 4 is a group-level version of the positivity assumption routinely made at the individual level in the analysis of observation studies. In the appendix, we show that the following theorem holds:

Theorem 6

Suppose that $f_{\mathbf{A}|\mathbf{L}i}(\cdot|\mathbf{L}_i)$ *satisfies assumptions 3 and 4, and that* α_0 *is the parametrization of a Bernoulli individual group assignment strategy (i.e. a type B parametrization) which satisfies assumption 4. Let* $\hat{Y}_i^{ipw}(a;\alpha_0) \equiv$

$$\frac{\sum_{j=1}^{n_i}\pi_i(\mathbf{A}_{i,-j};\alpha_0)1(A_{ij}=\alpha)Y_{ij}(\mathbf{A}_i)}{n_i \times f_{\mathbf{A}\left|\mathbf{L},i\right.}(\mathbf{A}_i\left|\mathbf{L}_i\right)}$$

and $\hat{Y}_i^{ipw}(\alpha_0) \equiv$

$$\frac{\sum_{j=1}^{n_i}\pi_i(\mathbf{A}_i;\alpha_0)Y_{ij}(\mathbf{A}_i)}{n_i \times f_{\mathbf{A}\left|\mathbf{L},i\right.}(\mathbf{A}_i\left|\mathbf{L}_i\right)}$$

Then

$$E\left\{\hat{Y}_i^{ipw}(a;\alpha_0)\right\}=\overline{Y}_i(a;\alpha_0)=\frac{1}{n_i}\sum_{j=1}^{n_i}\sum_{\mathbf{s}\in\mathscr{A}(n_i-1)}Y_{ij}(\mathbf{a}_{i,-j}=\mathbf{s},a_{ij}=a)\prod_{j'=1,j'\neq 1}^{n_i}\alpha_0^{s'_{ij}}(1-\alpha_0)^{1-s'_{ij}}$$

and

$$E\{\hat{Y}_i^{ipw}(\alpha_0)\}=\overline{Y}_i(\alpha_0)=\frac{1}{n_i}\sum_{j=1}^{n_i}\sum_{\mathbf{s}\in\mathscr{A}(n_i)}Y_{ij}(\mathbf{a}=\mathbf{s})\prod_{j'=1}^{n_i}\alpha_0^{s'_{ij}}(1-\alpha_0)^{1-s'_{ij}}$$

According to this theorem, if the allocation probability mechanism $f_{\mathbf{A}|\mathbf{L}}(\cdot|\mathbf{L}_i)$ is known, the population counterfactual averages $\overline{Y}_i(a;\alpha_0)$ and $\overline{Y}_i(\alpha_0)$ are identified from the observed data, and $\hat{Y}_i^{ipw}(a;\alpha_0)\text{and}\hat{Y}_i^{ipw}(\alpha_0)$ are unbiased estimators of $\overline{Y}_i(a;\alpha_0)$ and $\overline{Y}_i(\alpha_0)$ respectively. The theorem also immediately gives the following result. Let

$$\hat{DE}^{ipw}(\alpha_0)=\hat{Y}^{ipw}(0;\alpha_0)-\hat{Y}^{ipw}(1;\alpha_0),$$

$$\hat{IE}^{ipw}(\alpha_0,\alpha_1)=\hat{Y}^{ipw}(0;\alpha_0)-\hat{Y}^{ipw}(0;\alpha_1),$$

$$\hat{TE}^{ipw}(\alpha_0,\alpha_1)=\hat{Y}^{ipw}(0;\alpha_0)-\hat{Y}^{ipw}(0;\alpha_1),$$

$$\hat{OE}^{ipw}(\alpha_0,\alpha_1)=\hat{Y}^{ipw}(\alpha_0)-\hat{Y}^{ipw}(\alpha_1),$$

$$\text{Where}\hat{Y}^{ipw}(a;\alpha_0)=\sum_{i=1}^{N}\hat{Y}_i^{ipw}(a;\alpha_0)/N\text{and}\hat{Y}^{ipw}(\alpha_0)=\sum_{i=1}^{N}\hat{Y}_i^{ipw}(\alpha_0)/N.\text{Then,}$$

$$E\{\hat{DE}^{ipw}(\alpha_0)\}=\overline{DE}(\alpha_0),$$

$$E\{\hat{IE}^{ipw}(\alpha_0,\alpha_1)\}=\overline{IE}(\alpha_0,\alpha_1),$$

$$E\{\hat{TE}^{ipw}(\alpha_0,\alpha_1)\}=\overline{TE}(\alpha_0,\alpha_1),$$

$$E\{\hat{OE}^{ipw}(\alpha_0,\alpha_1)\}=\overline{OE}(\alpha_0,\alpha_1)$$

Unfortunately, $\hat{DE}^{ipw}(\alpha_0),\hat{IE}^{ipw}(\alpha_0,\alpha_1),\hat{TE}^{ipw}(\alpha_0,\alpha_1)\text{and}\hat{OE}^{ipw}(\alpha_0,\alpha_1)$ are not feasible in practice since, as is usually the case in observational studies, $f_{\mathbf{A}|\mathbf{L}}(\cdot|\mathbf{L}_i)$ is unknown to the analyst. To proceed, we must estimate this unknown treatment allocation mechanism from the observed data. Because $\mathbf{L}_i$ will typically include a large vector of

covariates, nonparametric estimation of $f_{\mathbf{A}|\mathbf{L}}(\mathbf{A}_i|\mathbf{L}_i)$ is not a viable option, and parametric or semi-parametric models must be adopted in practice. Next, we provide a brief and informal description to illustrate what a parametric approach entails in practice, in the particularly favorable setting where the number of groups N is reasonably large. In such a setting, we propose to estimate a parsimonious model $f_{\mathbf{A}|\mathbf{L}i}(\mathbf{A}_i|\mathbf{L}_i;\psi) = f_{\mathbf{A}|\mathbf{L}}(\mathbf{A}_i|\mathbf{L}_i;\psi)$ $i = 1, \ldots, N$, with unknown parameter $\psi = (\psi_a, \psi_b)$, where $f_{\mathbf{A}|\mathbf{L}}(\mathbf{A}_i|\mathbf{L}_i;\psi)$ is assumed to be a mixed model of the form

$$f_{\mathbf{A}\left|\mathbf{L}\right.}(\mathbf{A}_i \left|\mathbf{L}_i;\psi\right.) \equiv \int \prod_{j=1}^{n_i} h_{A\left|L\right.}(A_{ij}|L_{ij},\mathbf{b}_i;\psi_a) f_b(\mathbf{b}_i \left| V_i;\psi_b\right.) d\mathbf{b}_i$$

with $h_{A|L}(1|L_{ij}\,\mathbf{b}_i;\psi_a)$ say the logistic regression model logit $\{h_{A\left|L\right.}(1|L_{ij},\mathbf{b}_i;\psi_a)\} = \mathbf{b}_i + \psi_a' L_{ij}$ and $\mathbf{b}_i$ a random effect known to follow a parametric density $f_b(\mathbf{b}_i|V_i;\psi_b)$ indexed by an unknown parameter ψ_b. The standard logistic-normal mixed model corresponds to the choice of $f_b(\mathbf{b}_i|V_i;\psi_b)$ univariate normal with mean $\psi_{a,1}$ and variance $\psi_{a,2}$. Estimation of $\psi_a = (\psi_{a,1}, \psi_{a,2})$ and ψ_b is obtained by maximizing

$$\sum_{i=1}^{N} \log\{f_{\mathbf{A}\left|\mathbf{L}\right.}(\mathbf{A}_i|\mathbf{L}_i;\psi_a,\psi_b)\} \quad (18)$$

with respect to ψ to give $\hat{\psi}$. The mixed model paradigm is particularly appealing in the current setting, as it provides a flexible framework to account for a possible non-null conditional association between A_{ij} and $A_{ij'}$ given $\mathbf{L}_i$, for $j \neq j'$. Furthermore, under the assumption that $\mathbf{A}_i$ and $\mathbf{A}_{i'}$ are independent given $\mathbf{L}_i$ and $\mathbf{L}_{i'}$ for $i \neq i'$, $\hat{\psi}$ is a maximum likelihood estimator, and thus, under standard regularity conditions it is $\sqrt{N} - \text{consistent}$. However, note that the mixed model is agnostic to a possible non-null conditional association between A_{ij} and $A_{i'j'}$ for $i \neq i'$. Such a non-null association between the exposure levels of individuals belonging to different groups may arise say due to the spatial proximity of the two groups, even in the absence of between-group interference. In such a case, $\hat{\psi}$ is no longer the mle, but will remain consistent as the number of groups grows to infinity, provided that the non-null association of exposure levels between groups is not too pervasive. Specifically, this will hold provided that the dependence between the treatment allocation program of a given group is non-null only with that of a fixed number of groups, as determined say by spatial proximity. Feasible estimators of the various causal effects are then obtained by substituting $f_{\mathbf{A}\left|\mathbf{L}\right.}(\mathbf{A}_i \left|\mathbf{L}_i;\hat{\psi}\right.)$ for $f_{\mathbf{A}\left|\mathbf{L}_i\right.}(\mathbf{a}_i \left|\mathbf{L}_i\right.)$ in $\hat{Y}_i^{ipw}(a;\alpha_0)$ and $\hat{Y}_i^{ipw}(\alpha_0)$ Alternately, one may use the more stable estimators

$$\hat{Y}^{ipw}(a;\alpha_0,\hat{\psi}) = \frac{\sum_{i=1}^{N}[\sum_{j=1}^{n_i}\pi_i(\mathbf{A}_{i,-j};\alpha_0)1(A_{ij}=a)Yij(\mathbf{A}_i)/\{n_i \times f_{\mathbf{A}\left|\mathbf{L},i\right.}(\mathbf{A}_i|\mathbf{L}_{ii};\hat{\psi})\}]}{\sum_{i=1}^{N}[\sum_{j=1}^{n_i}\pi_i(\mathbf{A}_{i,-j};\alpha_0)1(A_{ij}=a)/\{n_i \times f_{\mathbf{A}\left|\mathbf{L},i\right.}(\mathbf{A}_i|\mathbf{L}_{ii};\hat{\psi})\}]}$$

$$\hat{Y}^{ipw}(\alpha_0,\hat{\psi})=\frac{\sum_{i=1}^{N}[\sum_{j=1}^{n_i}\pi_i(\mathbf{A}_i;\alpha_0)Y_{ij}(\mathbf{A}_i)/\{n_i\times f_{\mathbf{A}|\mathbf{L},i}(\mathbf{A}_i|\mathbf{L}_i;\hat{\psi})\}]}{\sum_{i=1}^{N}[\sum_{j=1}^{n_i}\pi_i(\mathbf{A}_i;\alpha_0)/\{n_i\times f_{\mathbf{A}|\mathbf{L},i}(\mathbf{A}_i|\mathbf{L}_i;\hat{\psi})\}]}$$

A large sample estimator of the variances of the estimates of the various causal effects can be obtained under standard regularity assumptions using well known Taylor series arguments that we do not reproduce here. The finite sample behavior of these various estimators will be examined in a simulation study we plan to report elsewhere.

Thus far we have assumed that $\mathbf{Y}_i(\cdot)$ is fixed; we will now briefly consider a setting in which $\mathbf{Y}_i(\cdot)$ is considered random. Hong and Raudenbush (2006) assume Stratified interference (Assumption 2) and assume that $Y_{ij}(\mathbf{a}_i)$ depends on $\mathbf{a}_{i,-j}$ only through some known scalar function $v(\mathbf{a}_{i,-j})$ so that $Y_{ij}(\mathbf{a}_i)$ can be written as $Y_{ij}(a_{ij}, v(\mathbf{a}_{i,-j}))$. Suppose now that for all i, j, A_{ij} is determined by simple randomization then assumption 3 will hold and it will also be the case that

$$E[Y_{ij}(a_{ij},v)|A_{ij},V(\mathbf{a}_{i,-j})]=E[Y_{ij}(a_{ij},v)]. \quad (19)$$

Hong and Raudenbush (2006) consider a variation on this assumption in the context of observational data. Specifically, they assume that

$$E[Y_{ij}(a_{ij},v)\,|A_{ij},V(\mathbf{a}_{i,-j}),L_{ij}]=E[Y_{ij}(a_{ij},v)|\,L_{ij}] \quad (20)$$

and from this it follows that

$$E[Y_{ij}(a,v)\,|L_{ij}=l_{ij}]=E[Y_{ij}|\,A_{ij}=a_{ij},V(\mathbf{a}_{i,-j})=v,L_{ij}=l_{ij}]$$

and from this one could obtain conditional direct, indirect and total effects, namely,

$$E[Y_{ij}(a,v)\,|L_{ij}=l_{ij}]-E[Y_{ij}(a',v)|\,L_{ij}=l_{ij}]$$

$$E[Y_{ij}(a,v)\,|L_{ij}=l_{ij}]-E[Y_{ij}(a,v')|\,L_{ij}=l_{ij}]$$

$$E[Y_{ij}(a,v)\,|L_{ij}=l_{ij}]-E[Y_{ij}(a',v')|\,L_{ij}=l_{ij}].$$

Hong and Raudenbush (2006) also allow L_{ij} to contain cluster level covariate along with cluster aggregates of individual level covariates. A similar approach is taken in VanderWeele (2010) in the context of mediation in the presence of interference. Note, however, that (20) requires that $Y_{ij}(a_{ij}, v)$ be mean independent of both A_{ij} and $V(\mathbf{a}_{i,-j})$ conditional on L_{ij}. If, for each individual A_{ij} is randomized conditional on L_{ij}, although this

will imply that $Y_{ij}(a_{ij}, v)$ is mean independent of A_{ij} conditional on L_{ij}, it does not necessarily guarantee that $Y_{ij}(a_{ij}, v)$ is mean independent of $V(\mathbf{a}_{i,-j}$ conditional on L_{ij}. More generally, instead of (21) we might consider

$$E[Y_{ij}(a_{ij}, v) \mid A_{ij}, V(\mathbf{a}_{i,-j}), L_{ij}, h(\mathbf{L}_i)] = E[Y_{ij}(a_{ij}, v) \mid L_{ij}, h(\mathbf{L}_i)] \quad (21)$$

where $h(\mathbf{L}_i)$ is a known function of $\mathbf{L}_i$. However once again, with (21), even if for each individual A_{ij} were randomized conditional on L_{ij}, $h(\mathbf{L}_i)$, this does not guarantee that $Y_{ij}(a_{ij}, v)$ is mean independent of $V(\mathbf{a}_{i,-j})$ conditional on L_{ij}, $h(\mathbf{L}_i)$ unless $h(\mathbf{L}_i) = \mathbf{L}_i$.

6 Varieties of direct and indirect effects

We have considered several types of effects that arise when there is interference between units. We have considered the effect on some outcome of an individual's treatment when the treatment of other units in a cluster are held fixed at a certain value; following, Hudgens and Halloran (2008), this was referred to as a"direct effect." We have also considered the effect on an individual's outcome of holding the individual's own treatment fixed but modifying the treatments received by other individuals in the same cluster; again following Hudgens and Halloran (2008), this was referred to as an "indirect effect." Of course, the terms "direct effects" and "indirect effects" are also used in the context of questions of mediation analysis, i.e. in assessing the extent to which the effect of some treatment on an outcome is mediated through some intermediate (the indirect effect) and the extent to which it occurs through other pathways (the direct effect). In some contexts, both interference and mediation may be present and of interest and the terms "direct effect" and "indirect effect" become ambiguous as they may make reference to the concepts from interference or from mediation.

In the infectious disease literature, the terminology of "direct and indirect effects" when interference is present dates at least as far back as Halloran and Struchiner (1991) although Hudgens and Halloran (2008) arguably provide the first formal counterfactual definitions. The terminology of "direct and indirect effects" in the context of mediation analysis extends at least as far back as the literature on structural equation modeling (e.g. Duncan, 1966) motivated by the method of path coefficients of Wright (1921); counterfactual notions of direct and indirect effects were described in detail by Holland (1988) and Robins and Greenland (1992). Because of the potential ambiguity in terms "direct effect" and "indirect effect," Sobel (2006) chose to use the term "spillover effect" for the effect on an individual's outcome of holding the individual's own treatment fixed but modifying the treatments received by other individuals. An early paper (Strain et al., 1976) in experimental educational psychology appears to have interchangeably used "indirect effect" and "spillover effect" to denote the effect on a child's outcome of holding the child's own treatment fixed but modifying the treatments received by other children. Complicating terminological issues yet further, the causal inference on mediation itself has produced alternative Definitions of direct and indirect effects based on potential interventions on the mediator (Robins and Greenland, 1992; Pearl, 2001) or alternatively on the notion of principal strata (Frangakis and Rubin, 2002; Rubin, 2004).

Variants of the notions of direct and indirect effects based on principal strata may in fact further be reformulated in the context of interference. Consider a vaccine trial (type A randomization) in which each cluster has two individuals so that for all i, $n_i = 2$ (e.g. a study of married households with no children) such that half of the households were randomized to no vaccine ($\alpha_0 = 0$) and half of the households were randomized to having one individual (e.g. the wife) vaccinated ($\alpha_1 = 0.5$). For each i, let $j = 1$ denote the subject that is potentially vaccinated (e.g. the wife) and $j = 2$ the subject that is never vaccinated (e.g. the husband). In the infectious disease context, a vaccination for individual 1 may prevent individual 2 from being infected either because the vaccine prevents individual 1 from being in infected or possibly because, even if individual 1 becomes infected, the vaccine itself renders the infection less contagious. A distinction between these two possibilities is sometimes drawn by using "susceptibility effect" to describe the former and "infectiousness effect" to describe the latter (Datta et al., 1999). Consider the following causal quantity, $E_i(Y_{i2}(1, 0) - Y_{i2}(0, 0)|Y_{i1}(1, 0) = Y_{i1}(0, 0) = 1)$; this is the effect on individual 2 of vaccinating individual 1 (with individual 2 unvaccinated) amongst the subset of households for whom individual 1 becomes infected irrespective of whether individual 1 receives the vaccination; this would be a principal strata direct effect (Rubin, 2004). If this quantity were non-zero we might interpret this as evidence of an "infectiousness effect" of the vaccine since the vaccination of individual 1 affects the outcome of individual 2 even though it has no effect on the outcome of individual 1. Future work could potentially adapt estimation methods for principal strata direct effects (Gallop et al, 2009; Sjölander et al., 2009) to attempt to estimate and potentially test for the presence of an "infectiousness effect", $E_i(Y_{i2}(1, 0) - Y_{i2}(0, 0)|Y_{i1}(1, 0) = Y_{i1}(0, 0) = 1)$.

Note that although the infectiousness effect quantity defined above is a "principal strata direct effect," within the context of interference it is a form of an "indirect effect" since individual 2's vaccination status is fixed to be unvaccinated in the causal comparison. Within the context of interference, both the "susceptibility effect" and the "infectiousness effect" are in fact forms of "indirect effects" (in the interference sense) because both the "susceptibility effect" and the "infectiousness effect" concern the effect on individual 2 of holding individual 2's vaccine status fixed but changing the vaccine status of individual 1; if interference were absent, neither of the effects would be present. If interference were absent then the principal strata "infectiousness effect" quantity defined above would reduce to $E_i(Y_{i2}(0) - Y_{i2}(0)|Y_{i1}(1) = Y_{i1}(0) = 1) = 0$. Again terminology concerning "direct and indirect effects" is ambiguous and is easily confused: what is a "direct effect" in the context of principal strata is an "indirect effect" in the context of interference.

Because of the multiple varieties of direct and indirect effects, the use of more specific terminology may be desirable. In the context of interference, "indirect effect" and "direct effect" could be replaced by "spillover effect" and "unit-treatment effect"; in the context of mediation, "indirect effect" and "direct effect" could be replaced by "mediated effect" and "unmediated effect." In the context of infectious diseases and the principal strata effect defined above, "susceptibility effect" and "infectiousness effect" could be used rather than making reference to "direct and indirect effects." Yet further caution with regard to

terminology on direct and indirect effects will be needed when both interference and mediation are present and of interest (VanderWeele, 2010).

7 Concluding remarks

In this paper we have reviewed some of the literature on causal inference in the presence of interference, we have provided new results on inference without the assumption of Stratified interference and we have described an inverse probability weighting approach to causal inference under interference in the context of observational studies. Interference arises in settings in which social interactions are present including settings of infectious disease, the study of neighborhoods and classrooms and in a variety of economic contexts. Although most work in causal inference has proceeded under a no-interference assumption, there are clearly many contexts in which such an assumption is not plausible. The issues raised by interference can be circumvented to a certain extent by implementing treatment programs at the cluster level rather than the individual level. However, interference gives rise to spillover effects which are themselves of intrinsic interest and the analysis of such spillover effects is inaccessible without explicitly taking interference into account. Theory and methods to address questions of interference and spillover effects will thus likely be important for a number of applied research settings.

The present work could be extended in a number of directions. Finite sample confidence intervals of shorter length than those in section 4 could be obtained by employing additional assumptions such as Stratified interference; continuous and unbounded outcomes could also be considered. The finite sample behavior of the inverse probability weighting estimation approach we proposed in this paper could be explored. Identification or partial Identification results for the "infectiousness effect," formalized in terms of principal strata, could be developed. Finally, further research could also potentially develop a more general framework for interference and spillover effects so as to consider a range of settings in which both interference and mediation were present and also so as to potentially allow for both within-cluster and between-cluster forms of interference. Causal inference under interference is a relatively new subfield and considerable work remains to be carried out.

References

1. Chow, YS.; Teicher, HP. Probability Theory: Independence, interchangeability, martingales. 3rd edition. Springer Texts in Statistics; 1997.
2. Datta S, Halloran ME, Longini IM. Efficiency of estimating vaccine efficacy for susceptibility and infectiousness: randomization by individual versus household. Biometrics. 1999; 55:792–798. [PubMed: 11315008]
3. Duncan OD. Path analysis: sociological examples. American Journal of Sociology. 1966; 72:1–16.
4. Frangakis CE, Rubin DB. Principal stratification in causal inference. Biometrics. 2002; 58:21–29. [PubMed: 11890317]
5. Gallop R, Small DS, Lin JY, Elliott MR, Joffe M, Ten Have TR. Mediation analysis with principal stratification. Statistics in Medicine. 2009; 28:1108–1130. [PubMed: 19184975]
6. Graham B. Identifying social interactions through conditional variance restrictions. Econometrica. 2008; 76:643–660.
7. Halloran ME, Struchiner CJ. Causal inference for infectious diseases. Epidemiology. 1995; 6:142–151. [PubMed: 7742400]

8. Hoeffding W. Probability inequalities for sums of bounded random variables. Journal of the American Statistical Association. 1963 Mar; 58(301):13–30.
9. Holland PW. Causal inference, path analysis, and recursive structural equations models. Sociological Methodology. 1988; 18:449–484.
10. Hong G, Raudenbush SW. Evaluating kindergarten retention policy: A case study of causal inference for multilevel observational data. Journal of the American Statistical Association. 2006; 101:901–910.
11. Hudgens MG, Halloran ME. Towards causal inference with interference. Journal of the American Statistical Association. 2008; 103:832–842. PMC2600548. [PubMed: 19081744]
12. Manski CF. Economic analysis of social interactions. Journal of Economic Perspectives. 2000; 14:115–136.
13. Manski CF. Identification of treatment response with social interactions. Northwestern University Working Paper. 2010
14. Joag-Dev K, Proschan F. Negative Association of Random Variables with Applications. Annals of Statistics. 1983; 11(1):286–295.
15. Pearl, J. Direct and indirect effects. Proceedings of the Seventeenth Conference on Uncertainty and Artificial Intelligence; San Francisco: Morgan Kaufmann. 2001. p. 411-420.
16. Robins JM, Greenland S. Identifiability and exchangeability for direct and indirect effects. Epidemiology. 1992; 3:143–155. [PubMed: 1576220]
17. Rosenbaum PR. Interference between units in randomized experiments. Journal of the American Statistical Association. 2007; 102:191–200.
18. Rubin DB. Comment on: "Randomization analysis of experimental data in the fisher randomization test" by D. Basu. Journal of the American Statistical Association. 1980; 75:591–593.
19. Rubin DB. Direct and indirect effects via potential outcomes. Scandinavian Journal of Statistics. 2004; 31:161–170.
20. Sjölander A, Humphreys K, Vansteelandt S, Bellocco R, Palmgren J. Sensitivity analysis for principal stratum direct effects, with an application to a study of physical activity and coronary heart disease. Biometrics. 2009; 65:514–520. [PubMed: 18759834]
21. Sobel ME. What Do Randomized Studies of Housing Mobility Demonstrate?: Causal Inference in the Face of Interference. Journal of the American Statistical Association. 2006; 101:1398–1407.
22. Strain PS, Shores RE, Kerr MM. An experimental analysis of "spillover" effects on the social interaction of behaviorally handicapped preschool children. Journal of Applied Behavior Analysis. 1976; 9:31–40. [PubMed: 1254540]
23. van der Vaart AW. Weak Convergence and Empirical Processes: With Applications to Statistics. Springer Serices in Statistics. 1996
24. VanderWeele TJ. Direct and indirect effects for neighborhood-based clustered and longitudinal data. Sociological Reserach and Methods. 2010 in press.
25. Wright S. Correlation and causation. J. Agric. Res. 1921; 20:557–585.

APPENDIX

Proof of Lemma 1

$$Var(\hat{Y}_i(1;\alpha_0)|S_i = 1) = Var\{\sum_{j=1}^{n_i} A_{ij}Y_{ij}(\mathbf{A}_i)/\sum_{j=1}^{n_i} A_{ij}\}$$

Note that

$$= \frac{1}{K_{0,i}^2}[\sum_i Var\{A_{ij}\, Y_{ij}(\mathbf{A}_i)\} + \sum_{j\neq j'} Cov\{A_{ij}\, Y_{ij}(\mathbf{A}_i), A_{ij'}\, Y_{ij'}(\mathbf{A}_i)\}]$$

Let $p_i \equiv \frac{1}{\binom{n_i}{K_i}}$. *Each term of the first sum equals*

$$Var\{A_{ij}\,Y_{ij}(\mathbf{A}_i)\}=Var\{\sum_{\omega\in\mathscr{A}(n_i-1,K_{0,i}-1)} 1(\mathbf{A}_{i,-j}=\omega)A_{ij}\,Y_{ij}(a_{ij}=1,\mathbf{a}_{i,-j}=\omega)\}$$

$$=\sum_{\omega\in\mathscr{A}(n_i-1,K_{0,i}-1)} Var\{1(\mathbf{A}_{i,-j}=\omega)A_{ij}\,Y_{ij}(a_{ij}=1,\mathbf{a}_{i,-j}=\omega)\}$$

$$+\sum_{(\omega,\omega')\in\mathscr{A}(n_i-1,K_{0,i}-1)} 1(\omega\neq\omega')\,Cov\left\{\begin{matrix}1(\mathbf{A}_{i,-j}=\omega)A_{ij}\,Y_{ij}(a_{ij}=1,\mathbf{a}_{i,-j}=\omega)\\ 1(\mathbf{A}_{i,-j}=\omega')A_{ij}\,Y_{ij}(a_{ij}=1,\mathbf{a}_{i,-j}=\omega')\end{matrix},\right\}$$

$$=p_i\{1-p_i\}\sum_{\omega\in\mathscr{A}(n_i-1,K_{0,i}-1)} Y_{ij}^2(a_{ij}=1,\mathbf{a}_{i,-j}=\omega)$$

$$-p_i^2\sum_{(\omega,\omega')\in\mathscr{A}(n_i-1,K_{0,i}-1)} 1(\omega\neq\omega')Y_{ij}(a_{ij}=1,\mathbf{a}_{i,-j}=\omega)Y_{ij}(a_{ij}=1,\mathbf{a}_{i,-j}=\omega')$$

and

$$Cov\{A_{ij}\,Y_{ij}(\mathbf{A}_i),A_{ij'}\,Y_{ij'}(\mathbf{A}_i)\}=E\{A_{ij}\,Y_{ij}(\mathbf{A}_i)A_{ij'}\,Y_{ij'}(\mathbf{A}_i)\}-E\{A_{ij}\,Y_{ij}(\mathbf{A}_i)\}E\{A_{ij'}\,Y_{ij'}(\mathbf{A}_i)\}$$

$$=\sum_{\omega\in\mathscr{A}(n_i-2,K_{0,i}-2)} p_iY_{ij}(a_{ij}=1,a_{ij'}=1,\mathbf{a}_{i,-(j,j')}=\omega)Y_{ij'}(a_{ij}=1,a_{ij'}=1,\mathbf{a}_{i,-(j,j')}=\omega)$$

$$-\sum_{(\omega,\omega')\in\mathscr{A}(n_i-1,K_{0,i}-1)} p_i^2Y_{ij}(a_{ij}=1,\mathbf{a}_{i,-j}=\omega)Y_{ij'}(a_{ij'}=1,\mathbf{a}_{i,-j'}=\omega')$$

$$=\sum_{\omega\in\mathscr{A}(n_i-2,K_{0,i}-2)} p_iY_{ij}(a_{ij}=1,a_{ij'}=1,\mathbf{a}_{i,-(j,j')}=\omega)Y_{ij'}(a_{ij}=1,a_{ij'}=1,\mathbf{a}_{i,-(j,j')}=\omega)$$

$$-\sum_{\omega\in\mathscr{A}(n_i-2,K_{0,i}-2)} p_i^2Y_{ij}(a_{ij}=1,a_{ij'}=1,\mathbf{a}_{i,-(j,j')}=\omega)Y_{ij'}(a_{ij}=1,a_{ij'}=1,\mathbf{a}_{i,-(j,j')}=\omega)$$

$$-\sum_{(\omega,\omega')\in\mathscr{A}(n_i-1,K_{0,i}-1)} 1(\omega\neq\omega')p_i^2Y_{ij}(a_{ij}=1,\mathbf{a}_{i,-j}=\omega)Y_{ij'}(a_{ij'}=1,\mathbf{a}_{i,-j'}=\omega')$$

so that

$$Var(\hat{Y}_i(1;\alpha_0)|S_i=1)=\frac{1}{K_{0,i}^2}p_i\{1-p_i\}$$

$$\times\left[\begin{array}{c}\Sigma_j \sum_{\omega\in\mathscr{A}(n_i-1,K_{0,i}-1)} Y_{ij}^2(a_{ij}=1,\mathbf{a}_{i,-j}=\omega)\\ +\Sigma_{j\neq j'}\sum_{\omega\in\mathscr{A}(n_i-2,K_{0,i}-2)} Y_{ij}(a_{ij}=1,a_{ij'}=1,\mathbf{a}_{i,-(j,j')}=\omega)Y_{ij'}(a_{ij}=1,a_{ij'}=1,\mathbf{a}_{i,-(j,j')}=\omega)\end{array}\right]$$

$$-\frac{1}{K_{0,i}^2}p_i^2\left[\begin{array}{c}\Sigma_j \sum_{(\omega,\omega')\in\mathscr{A}(n_i-1,K_{0,i}-1)} 1(\omega\neq\omega')Y_{ij}(a_{ij}=1,\mathbf{a}_{i,-j}=\omega)Y_{ij}(a_{ij}=1,\mathbf{a}_{i,-j}=\omega')\\ +\Sigma_{j\neq j'}\sum_{(\omega,\omega')\in\mathscr{A}(n_i-1,K_{0,i}-1)} 1(\omega\neq\omega')Y_{ij}(a_{ij}=1,\mathbf{a}_{i,-j}=\omega)Y_{ij'}(a_{ij'}=1,\mathbf{a}_{i,-j'}=\omega')\end{array}\right]$$

$$=\frac{1}{K_{0,i}^2}p_i\{1-p_i\}$$

$$\times\left[\begin{array}{c}\Sigma_j \sum_{\omega\in\mathscr{A}(n_i-1,K_{0,i}-1)} Y_{ij}^2(a_{ij}=1,\mathbf{a}_{i,-j}=\omega)\\ +\Sigma_{j\neq j'}\sum_{\omega\in\mathscr{A}(n_i-2,K_{0,i}-2)} Y_{ij}(a_{ij}=1,a_{ij'}=1,\mathbf{a}_{i,-(j,j')}=\omega)Y_{ij'}(a_{ij}=1,a_{ij'}=1,\mathbf{a}_{i,-(j,j')}=\omega)\end{array}\right]$$

$$-\frac{1}{K_{0,i}^2}p_i^2\left[\Sigma_{j,j'}\sum_{(\omega,\omega')\in\mathscr{A}(n_i-1,K_{0,i}-1)} 1(\omega\neq\omega')Y_{ij}(a_{ij}=1,\mathbf{a}_{i,-j}=\omega)Y_{ij'}(a_{ij}=1,\mathbf{a}_{i,-j}=\omega')\right]$$

Therefore, as $Y_{ij}(\mathbf{a}_i)\geq 0$ *for all* $\mathbf{a}_i\in A(n_i;K_{0,i};)$ *and all j in group i,*

$E\{\hat{Var}_u(\hat{Y}_i(1;\alpha_0)\,|S_i=1)|S_i=1\}>Var(\hat{Y}_i(1;\alpha_0)\,|\,S_i=1)$, since

$$E\{\hat{Var}_u(\hat{Y}_i(1;\alpha_0)\,|S_i=1)|\,S_i=1\}$$

$$=\frac{1}{K_{0,i}^2}p_i\{1-p_i\}$$

$$\times\left[\begin{array}{c}\Sigma_j \sum_{\omega\in\mathscr{A}(n_i-1,K_{0,i}-1)} Y_{ij}^2(a_{ij}=1,\mathbf{a}_{i,-j}=\omega)\\ +\Sigma_{j\neq j'}\sum_{\omega\in\mathscr{A}(n_i-2,K_{0,i}-2)} Y_{ij}(a_{ij}=1,a_{ij'}=1,\mathbf{a}_{i,-(j,j')}=\omega)Y_{ij'}(a_{ij}=1,a_{ij'}=1,\mathbf{a}_{i,-(j,j')}=\omega)\end{array}\right].$$

Proof of Theorems 1-5

See technical report available from the authors.

Proof of Theorem 6

Under Assumptions 3 and 4, we have that for $a=0,1; E\{\hat{Y}_i^{ipw}(a;\alpha_0)\}$

$$=\frac{1}{n_i}\sum_{\mathbf{s}\in\mathscr{A}(n_i)}\frac{\Pr\{\mathbf{A}_i=\mathbf{s}|\mathbf{L}_i,\boldsymbol{Y}_i(\cdot)\}}{f_{\mathbf{A}|\mathbf{L},i}(\mathbf{s}|\mathbf{L}_i)}\sum_{j=1}^{n_i}1(s_{ij}=a)Y_{ij}(\mathbf{a}_{i,-j}=\mathbf{s},a_{ij}=s_{ij})\prod_{j'=1,j'\neq j}^{n_i}\alpha_0^{s_{ij'}}(1-\alpha_0)^{1-s_{ij'}}$$

$$=\frac{1}{n_i}\sum_{\mathbf{s}\in\mathscr{A}(n_i)}\frac{f_{\mathbf{A}|\mathbf{L},i}(\mathbf{s}|\mathbf{L}_i)}{f_{\mathbf{A}|\mathbf{L},i}(\mathbf{s}|\mathbf{L}_i)}\sum_{j=1}^{n_i}1(s_{ij}=a)Y_{ij}(\mathbf{a}_{i,-j}=\mathbf{s},a_{ij}=s_{ij})\prod_{j'=1,j'\neq j}^{n_i}\alpha_0^{s_{ij'}}(1-\alpha_0)^{1-s_{ij'}}$$

$$=\frac{1}{n_i}\sum_{j=1}^{n_i}\sum_{\mathbf{s}\in\mathscr{A}(n_i-1)}Y_{ij}(\mathbf{a}_{i,-j}=\mathbf{s},a_{ij}=a)\prod_{j'=1,j'\neq j}^{n_i}\alpha_0^{s_{ij'}}(1-\alpha_0)^{1-s_{ij'}}$$

similarly, $E\{\hat{Y}_i^{ipw}(\alpha_0)\}$

$$=\frac{1}{n_i}\sum_{\mathbf{s}\in\mathscr{A}(n_i)}\frac{\Pr\{\mathbf{A}_i=\mathbf{s}|\mathbf{L}_i,\boldsymbol{Y}_i(\cdot)\}}{f_{\mathbf{A}|\mathbf{L},i}(\mathbf{s}|\mathbf{L}_i)}\sum_{j=1}^{n_i}Y_{ij}(\mathbf{a}_{ij}=\mathbf{s})\prod_{j'=1,}^{n_i}\alpha_0^{s_{ij'}}(1-\alpha_0)^{1-s_{ij'}}\overline{Y}_i(a;\alpha_0)$$

$$=\frac{1}{n_i}\sum_{\mathbf{s}\in\mathscr{A}(n_i)}\frac{f_{\mathbf{A}|\mathbf{L},i}(\mathbf{s}|\mathbf{L}_i)}{f_{\mathbf{A}|\mathbf{L},i}(\mathbf{s}|\mathbf{L}_i)}\sum_{j=1}^{n_i}Y_{ij}(\mathbf{a}_{i,j}=\mathbf{s})\prod_{j'=1,}^{n_i}\alpha_0^{s_{ij'}}(1-\alpha_0)^{1-s_{ij'}}$$

$$=\frac{1}{n_i}\sum_{j=1}^{n_i}\sum_{\mathbf{s}\in\mathscr{A}(n_i)}Y_{ij}(\mathbf{a}=\mathbf{s})\prod_{j'=1}^{n_i}\alpha_0^{A'_{ij}}(1-\alpha_0)^{1-A_{ij'}}$$