

# On inverse probability-weighted estimators in the presence of interference

BY L. LIU

*School of Statistics, University of Minnesota at Twin Cities, 224 Church St SE #313,  
Minneapolis, Minnesota 55455, U.S.A.*

liux3771@umn.edu

M. G. HUDGENS

*Department of Biostatistics, University of North Carolina, CB #7420, Chapel Hill,  
North Carolina 27599, U.S.A.*

mhudgens@email.unc.edu

AND S. BECKER-DREPS

*Department of Family Medicine, University of North Carolina, CB #7595, Chapel Hill,  
North Carolina 27599, U.S.A.*

sbd@email.unc.edu

## SUMMARY

We consider inference about the causal effect of a treatment or exposure in the presence of interference, i.e., when one individual's treatment affects the outcome of another individual. In the observational setting where the treatment assignment mechanism is not known, inverse probability-weighted estimators have been proposed when individuals can be partitioned into groups such that there is no interference between individuals in different groups. Unfortunately this assumption, which is sometimes referred to as partial interference, may not hold, and moreover existing weighted estimators may have large variances. In this paper we consider weighted estimators that could be employed when interference is present. We first propose a generalized inverse probability-weighted estimator and two Hájek-type stabilized weighted estimators that allow any form of interference. We derive their asymptotic distributions and propose consistent variance estimators assuming partial interference. Empirical results show that one of the Hájek estimators can have substantially smaller finite-sample variance than the other estimators. The different estimators are illustrated using data on the effects of rotavirus vaccination in Nicaragua.

*Some key words:* Causal inference; Interference; Inverse probability-weighted estimator; Observational study.

## 1. INTRODUCTION

In causal inference it is often assumed that there is no interference between individuals, i.e., that the treatment of one individual does not affect the outcome of another. However, this assumption may not hold. For instance, in infectious disease studies, the vaccination status of one individual may affect whether another individual becomes infected (Halloran & Struchiner, 1995). Similarly, encouraging one individual to vote may increase the likelihood that another individual in

the same household will vote (Nickerson, 2008). Interference may also occur between students in the same classroom (Hong & Raudenbush, 2006) or between households in the same neighbourhood (Sobel, 2006), and in myriad other contexts (Rosenbaum, 2007; Luo et al., 2012; Manski, 2013).

Inference in the presence of interference is interesting, because a treatment may have multiple types of effects, but difficult, because individuals may have many potential outcomes. Recently, methods have been developed for the setting where individuals can be partitioned into groups such that there may be interference between individuals in the same group but not between individuals in different groups; this is sometimes called partial interference (Sobel, 2006). Assuming partial interference, Hudgens & Halloran (2008) defined the direct, indirect, total and overall causal effects of a treatment in randomized studies. Inference about these types of causal effects has subsequently been considered by VanderWeele & Tchetgen Tchetgen (2011), VanderWeele et al. (2012), Halloran & Hudgens (2012), Liu & Hudgens (2014) and P. M. Aronow and C. Samii in an unpublished 2013 paper (arXiv:1305.6156), among others. For observational settings where the treatment assignment mechanism is not known, Tchetgen Tchetgen & VanderWeele (2012) proposed inverse probability-weighted estimators of these causal effects based on group-level propensity scores. These weighted estimators can be viewed as a generalization of the usual inverse probability-weighted estimator of the causal effect of a treatment in the absence of interference. However, in general, weighted estimators are known to have relatively large variance. Additionally, in some settings the partial interference assumption may be dubious. In this article we consider alternative weighted-type estimators that allow for general forms of interference and tend to be less variable.

## 2. PRELIMINARIES

Consider a finite population of  $n$  individuals, and suppose that each individual may receive some treatment or exposure. Let  $Z_i$  ( $i = 1, \dots, n$ ) be the random variable such that  $Z_i = 1$  if individual  $i$  received treatment and  $Z_i = 0$  otherwise. Suppose that interference may be present between the  $n$  individuals, and define the interference set  $\chi_i = \{i_1, i_2, \dots\}$  for individual  $i$  to be an ordered set of all other individuals whose treatment received might affect the outcome of individual  $i$ . Assume that there is no interference between individual  $i$  and individuals not in  $\chi_i$ . There may or may not be interference between individual  $i$  and individuals in  $\chi_i$ . A central goal of the inferential methods described below is to quantify the extent to which such interference is present. Let  $S_i = (Z_{i_1}, Z_{i_2}, \dots)$  denote the vector of treatment indicators for individuals that possibly interfere with individual  $i$ ; that is, the outcome of individual  $i$  is allowed to depend not only on  $Z_i$  but also on  $S_i$ . For example, if the outcome of individual 1 possibly depends on their own treatment status as well as that of individuals 2 and 3 but not on that of individuals 4,  $\dots, n$ , then  $\chi_1 = \{2, 3\}$  and  $S_1 = (Z_2, Z_3)$ . The interference sets  $\chi_1, \dots, \chi_n$  are assumed to be known a priori. Denote possible values of  $Z_i$  and  $S_i$  by  $z_i$  and  $s_i$ . Let  $y_i(z_i, s_i)$  denote the potential outcome of individual  $i$  if they receive treatment  $z_i$  and their interference set receives  $s_i$ . This potential outcome notation is general enough to encompass any possible interference structure, of which partial interference is a special case. Let  $Y_i = y_i(Z_i, S_i)$  denote the observed outcome. The potential outcomes  $y_i(z_i, s_i)$  are assumed to be deterministic functions of  $z_i$  and  $s_i$ , and the observed outcome  $Y_i$  is considered to be random because it depends on the random variables  $Z_i$  and  $S_i$ . Let  $\sum S_i$  be the sum over all the components of  $S_i$ , and let  $|S_i|$  denote the dimension of the vector  $S_i$ . For example, if  $S_1 = (Z_2, Z_3)$ , then  $\sum S_1 = Z_2 + Z_3$  and  $|S_1| = 2$ .

In the absence of interference, a common causal estimand is the average treatment effect, which contrasts the average outcome for the counterfactual scenario where every individual

in the population is treated with that of the counterfactual scenario where every individual in the population is not treated. Similarly, in the presence of interference, causal estimands can be defined in terms of counterfactual scenarios corresponding to different treatment allocation strategies (e.g., Hong & Raudenbush, 2006; Sobel, 2006; Hudgens & Halloran, 2008; Tchetgen Tchetgen & VanderWeele, 2012). For example, the indirect effect, defined formally below, contrasts average outcomes of untreated individuals for the counterfactual scenario where one allocation strategy is adopted in the population with those for the counterfactual scenario where some other allocation strategy is adopted in the population. Such estimands quantify interference, if present, at the population level and can be used to inform policy decisions regarding a treatment or exposure. The allocation strategy of interest will in general depend on the setting.

Here we consider Bernoulli allocation strategies proposed by Tchetgen Tchetgen & VanderWeele (2012), where strategy  $\alpha$  corresponds to the counterfactual scenario in which individuals independently receive treatment with probability  $\alpha$ . It is not assumed that the observed treatment indicators  $Z_1, \dots, Z_n$  are independent Bernoulli random variables; rather, the distribution of treatment under Bernoulli allocation is used below to define the counterfactual estimands of interest. By analogy, direct standardization of mortality rates could entail using the 2010 United States census age distribution, which may differ from the age distribution giving rise to the observed data. Corresponding to Bernoulli allocation, let  $\pi(s_i; \alpha) = \alpha^{\sum s_i} (1 - \alpha)^{|S_i| - \sum s_i}$  denote the probability of the interference set for individual  $i$  receiving treatment  $s_i$  under allocation strategy  $\alpha$ . Let  $\pi(z_i; \alpha) = \alpha^{z_i} (1 - \alpha)^{1 - z_i}$  and  $\pi(z_i, s_i; \alpha) = \pi(z_i; \alpha) \pi(s_i; \alpha)$  denote, respectively, the probability of individual  $i$  receiving treatment  $z_i$  and the probability of individual  $i$  together with their interference set receiving joint treatment  $(z_i, s_i)$  under allocation strategy  $\alpha$ . Define  $\bar{y}_i(z, \alpha) = \sum_{s_i} y_i(z_i = z, s_i) \pi(s_i; \alpha)$  to be the average potential outcome of individual  $i$  under allocation strategy  $\alpha$ , where the summation is over all  $2^{|S_i|}$  possible values of  $s_i$ . Returning to the example where  $S_1 = (Z_2, Z_3)$ , the average potential outcome of individual 1 is a weighted average of potential outcomes under different combinations of treatment  $Z_1 = z$  and  $(Z_2, Z_3) \in \{(0, 0), (0, 1), (1, 0), (1, 1)\}$ , with the weights being the corresponding probabilities under Bernoulli allocation. Averaging over all individuals, define the population average potential outcome as  $\bar{y}(z, \alpha) = \sum_{i=1}^n \bar{y}_i(z, \alpha) / n$ . Similarly, define the marginal average potential outcome for individual  $i$  under allocation strategy  $\alpha$  by  $\bar{y}_i(\alpha) = \sum_{z_i, s_i} y_i(z_i, s_i) \pi_i(z_i, s_i; \alpha)$  and define the population marginal average potential outcome as  $\bar{y}(\alpha) = \sum_{i=1}^n \bar{y}_i(\alpha) / n$ .

Various causal effects can be defined by contrasts in the population average potential outcomes. In particular, define the direct effect of treatment under allocation strategy  $\alpha$  to be  $\bar{D}\bar{E}(\alpha) = g\{\bar{y}(1, \alpha), \bar{y}(0, \alpha)\}$ , where  $g(\cdot, \cdot)$  is some continuous contrast function. A commonly used contrast function is  $g(x_1, x_0) = x_1 - x_0$ ; in vaccine trials with a binary outcome it is typical to use  $g(x_1, x_0) = 1 - x_1/x_0$ . The direct effect compares the average potential outcomes when an individual receives treatment versus not under allocation strategy  $\alpha$ . For two allocation strategies  $\alpha_1$  and  $\alpha_0$ , let  $\bar{I}\bar{E}(\alpha_1, \alpha_0) = g\{\bar{y}(0, \alpha_1), \bar{y}(0, \alpha_0)\}$  be the indirect or spillover effect, which contrasts average potential outcomes when individuals do not receive treatment under different allocation strategies. In the context of vaccines, the indirect effect is sometimes referred to as herd immunity and describes the effect of the proportion of individuals vaccinated, e.g., 30% versus 50%, on the average outcome among unvaccinated individuals. An indirect effect can also be defined for when individuals receive treatment,  $z = 1$ , but for simplicity we do not consider such indirect effects here. The total effect  $\bar{T}\bar{E}(\alpha_1, \alpha_0) = g\{\bar{y}(1, \alpha_1), \bar{y}(0, \alpha_0)\}$  incorporates both direct and indirect effects, and reflects the difference between the average potential outcomes when individuals receive treatment under one allocation strategy versus when they go without treatment under another allocation strategy. Finally, define  $\bar{O}\bar{E}(\alpha_1, \alpha_0) = g\{\bar{y}(\alpha_1), \bar{y}(\alpha_0)\}$  to be the overall

effect, which describes the contrast in average outcomes under one allocation strategy relative to another.

3. INVERSE PROBABILITY-WEIGHTED AND HÁJEK-TYPE ESTIMATORS

In this section we propose inverse probability-weighted and Hájek-type estimators which allow for general interference; that is, no assumption is made regarding the structure or form of interference that might be present. When there is partial interference and the groups are of the same size, the inverse probability-weighted estimators defined below reduce to those proposed by Tchetgen Tchetgen & VanderWeele (2012). Aronow and Samii (arXiv:1305.6156) considered similar estimators in the setting where interference may be present, but where treatment is assigned randomly according to a known experimental design.

Let  $l_i$  denote a vector of pretreatment covariates of individual  $i$ , and let  $l_{\chi_i} = (l_{i_1}, l_{i_2}, \dots)$ . Assume that conditional on covariates  $l_i$ , the treatment allocation for individual  $i$  is independent of all potential outcomes and other covariates; that is,  $\text{pr}(Z_i = z_i \mid l_i) = \text{pr}\{Z_i = z_i \mid l_1, \dots, l_n, y_1(\cdot), \dots, y_n(\cdot)\}$ . Likewise, assume  $\text{pr}(Z_i = z_i, S_i = s_i \mid l_i, l_{\chi_i}) = \text{pr}\{Z_i = z_i, S_i = s_i \mid l_1, \dots, l_n, y_1(\cdot), \dots, y_n(\cdot)\}$ . Define  $f(z_i \mid l_i) = \text{pr}(Z_i = z_i \mid l_i)$  and  $f(z_i, s_i \mid l_i, l_{\chi_i}) = \text{pr}(Z_i = z_i, S_i = s_i \mid l_i, l_{\chi_i})$  to be the propensity scores of individual  $i$  and of individual  $i$  and their interference set, respectively. Assume that  $f(z_i \mid l_i) > 0$  and  $f(z_i, s_i \mid l_i, l_{\chi_i}) > 0$  for all  $z_i, s_i, l_i$  and  $l_{\chi_i}$ . Define the inverse probability-weighted estimator for treatment  $z$  under allocation strategy  $\alpha$  to be

$$\hat{Y}^{\text{ipw}}(z, \alpha) = n^{-1} \sum_i \frac{y_i(Z_i, S_i) 1(Z_i = z) \pi(S_i; \alpha)}{f(Z_i, S_i \mid l_i, l_{\chi_i})} \quad (z = 0, 1), \tag{1}$$

and define the inverse probability-weighted marginal estimator under strategy  $\alpha$  to be

$$\hat{Y}^{\text{ipw}}(\alpha) = n^{-1} \sum_i \frac{y_i(Z_i, S_i) \pi(Z_i, S_i; \alpha)}{f(Z_i, S_i \mid l_i, l_{\chi_i})}, \tag{2}$$

where  $\sum_i$  means  $\sum_{i=1}^n$ . If the propensity scores are known, then (1) and (2) are unbiased as stated in the following proposition.

PROPOSITION 1. *If  $f(Z_i, S_i \mid l_i, l_{\chi_i})$  is known for all  $i$ , then  $E\{\hat{Y}^{\text{ipw}}(z, \alpha)\} = \bar{y}(z, \alpha)$  and  $E\{\hat{Y}^{\text{ipw}}(\alpha)\} = \bar{y}(\alpha)$ .*

In the absence of interference, the Hájek (1971) estimator of the mean of a finite population replaces the denominator  $n$  of the Horvitz & Thompson (1952) inverse probability-weighted estimator with the sum of the inverse of the sampling probabilities, which tends to reduce the variance relative to the Horvitz–Thompson estimator. Returning to the current context, let  $\hat{n}_{1z} = \sum_i 1(Z_i = z) / f(Z_i \mid l_i)$  and note that  $E(\hat{n}_{1z}) = n$  even if interference is present. This suggests replacing  $n$  in (1) with  $\hat{n}_{1z}$  to obtain a stabilized Hájek-type estimator. Alternatively, notice that the weighted estimator (1) involves  $f(Z_i, S_i \mid l_i, l_{\chi_i})$ , which suggests replacing  $n$  with the unbiased estimator  $\hat{n}_{2z} = \sum_i 1(Z_i = z) \pi(S_i; \alpha) / f(Z_i, S_i \mid l_i, l_{\chi_i})$  instead. Therefore, we will consider two different Hájek-type estimators of the population average outcome for treatment  $z$  and allocation strategy  $\alpha$ , defined by

$$\hat{Y}_h^{\text{haj}}(z, \alpha) = \hat{n}_{hz}^{-1} \sum_i \frac{y_i(Z_i, S_i) 1(Z_i = z) \pi(S_i; \alpha)}{f(Z_i, S_i \mid l_i, l_{\chi_i})} \quad (h = 1, 2).$$

Here and below we assume that there exists at least one  $i$  such that  $Z_i = z$  for  $z = 0, 1$ . Similarly, let  $\hat{n}_1 = \sum_i \pi(Z_i; \alpha) / f(Z_i | l_i)$  and  $\hat{n}_2 = \sum_i \pi(Z_i, S_i; \alpha) / f(Z_i, S_i | l_i, l_{\chi_i})$ , and note that  $E(\hat{n}_h) = n$  ( $h = 1, 2$ ), which suggests the following estimators of the population average marginal outcome for allocation strategy  $\alpha$ :

$$\hat{Y}_h^{\text{haj}}(\alpha) = \hat{n}_h^{-1} \sum_i \frac{y_i(Z_i, S_i) \pi(Z_i, S_i; \alpha)}{f(Z_i, S_i | l_i, l_{\chi_i})} \quad (h = 1, 2).$$

Note that  $\hat{n}_{2z}$ ,  $\hat{n}_1$  and  $\hat{n}_2$  depend on  $\alpha$ , but we suppress this dependence for notational convenience. In what follows,  $\hat{Y}_1^{\text{haj}}(\cdot)$  and  $\hat{Y}_2^{\text{haj}}(\cdot)$  will be referred to as the Hájek 1 and Hájek 2 estimators.

An appealing property of  $\hat{Y}_2^{\text{haj}}(z, \alpha)$  and  $\hat{Y}_2^{\text{haj}}(\alpha)$  is the preservation of the bounds of the potential outcome  $y_i(\cdot)$ . Specifically, suppose there exist constants  $m_l$  and  $m_u$  such that  $m_l \leq y_i(\cdot) \leq m_u$  ( $i = 1, \dots, n$ ); then  $m_l \leq \hat{Y}_2^{\text{haj}}(z, \alpha) \leq m_u$  and  $m_l \leq \hat{Y}_2^{\text{haj}}(\alpha) \leq m_u$ . For example, if  $y_i(\cdot)$  is binary, then  $\hat{Y}_2^{\text{haj}}(z, \alpha), \hat{Y}_2^{\text{haj}}(\alpha) \in [0, 1]$ . In contrast, preservation of the bounds is not guaranteed for  $\hat{Y}^{\text{ipw}}(\cdot)$  or  $\hat{Y}_1^{\text{haj}}(\cdot)$ .

Another attractive property of the Hájek 2 estimators is preservation of linear transformations of the outcome. In particular, suppose that the observed outcomes  $Y_i$  are transformed by the function  $\mathcal{L}(x) = ax + b$  ( $a, b \in \mathbb{R}$ ). Then Hájek 2 estimators computed using the transformed responses will equal  $\mathcal{L}\{\hat{Y}_2^{\text{haj}}(z, \alpha)\}$  and  $\mathcal{L}\{\hat{Y}_2^{\text{haj}}(\alpha)\}$ , where  $\hat{Y}_2^{\text{haj}}(z, \alpha)$  and  $\hat{Y}_2^{\text{haj}}(\alpha)$  are computed on the original, untransformed observed outcomes. In contrast, the inverse probability-weighted and Hájek 1 estimators have this property only when  $b = 0$ .

Define  $\hat{DE}^{\text{ipw}}(\alpha) = g\{\hat{Y}^{\text{ipw}}(1, \alpha), \hat{Y}^{\text{ipw}}(0, \alpha)\}$  to be the inverse probability-weighted estimator of the direct effect. Define  $\hat{IE}^{\text{ipw}}(\alpha_1, \alpha_0) = g\{\hat{Y}^{\text{ipw}}(0, \alpha_1), \hat{Y}^{\text{ipw}}(0, \alpha_0)\}$ ,  $\hat{TE}^{\text{ipw}}(\alpha_1, \alpha_0) = g\{\hat{Y}^{\text{ipw}}(1, \alpha_1), \hat{Y}^{\text{ipw}}(0, \alpha_0)\}$  and  $\hat{OE}^{\text{ipw}}(\alpha_1, \alpha_0) = g\{\hat{Y}^{\text{ipw}}(\alpha_1), \hat{Y}^{\text{ipw}}(\alpha_0)\}$  to be the weighted estimators of the indirect, total and overall effects. Hájek-type causal effect estimators are defined similarly. For example, define Hájek-type estimators of the direct effect by  $\hat{DE}_h^{\text{haj}}(\alpha) = g\{\hat{Y}_h^{\text{haj}}(1, \alpha), \hat{Y}_h^{\text{haj}}(0, \alpha)\}$  ( $h = 1, 2$ ). If the contrast function is  $g(x_1, x_0) = x_1 - x_0$ , then by the property described in the preceding paragraph, the values of Hájek 2 causal effect estimators are invariant under location shift. This is not the case for the inverse probability-weighted and Hájek 1 causal effect estimators.

#### 4. ASYMPTOTIC DISTRIBUTIONS

In this section the large-sample properties of the inverse probability-weighted and Hájek-type estimators are derived assuming partial interference. In particular, assume that individuals can be partitioned into groups such that there is no interference between individuals in different groups. Within groups no additional structure is assumed regarding interference, so there may be interference between any two individuals within a group. That is, we assume the following.

*Assumption 1.* There exists a partition  $\{C_\nu\}_{\nu=1}^m$  of  $\{1, \dots, n\}$  such that  $\chi_i = C_\nu \setminus \{i\}$  ( $i \in C_\nu$ ;  $\nu = 1, \dots, m$ ).

Let  $N_\nu = |C_\nu|$  denote the number of individuals in group  $\nu$ . Let  $Y_{vi}$  denote the observed outcome for individual  $i$  in group  $\nu$ , and write  $\tilde{Y}_\nu = (Y_{\nu 1}, \dots, Y_{\nu N_\nu})$ . Let  $L_{vi}$  and  $Z_{vi}$  denote the observed covariates and treatment for individual  $i$  in group  $\nu$ , and define  $\tilde{L}_\nu$  and  $\tilde{Z}_\nu$  analogously to  $\tilde{Y}_\nu$ . Assume that  $N_\nu$  is one of the baseline covariates included in  $L_{vi}$ .

To derive the large-sample properties of the inverse probability-weighted and Hájek-type estimators, assume that the  $m$  groups are a random sample from an infinite superpopulation of groups such that the observable random variables  $(\tilde{Y}_v, \tilde{Z}_v, \tilde{L}_v)$  ( $v = 1, \dots, m$ ) are independent and identically distributed. Let  $F$  denote the distribution function of  $(\tilde{Y}_v, \tilde{Z}_v, \tilde{L}_v)$ .

Let  $Y_{vi}(z, s)$  denote the potential outcome for individual  $i$  in group  $v$ , where  $z$  denotes treatment received by individual  $i$  and  $s$  denotes the vector of treatment indicators for all other individuals in group  $v$ . Unlike in § 2 and 3, here the potential outcomes are considered random variables because of the assumed random sampling of the  $m$  groups from a superpopulation. Denote the observed outcome for individual  $i$  by  $Y_{vi} = Y_{vi}(Z_{vi}, S_{vi})$ , where  $S_{vi}$  is the subvector of  $\tilde{Z}_v$  with  $Z_{vi}$  removed. Note that  $S_{vi}$  is a function of  $\tilde{Z}_v$ , which for notational simplicity is left implicit. Assume conditional exchangeability, i.e.,  $Y_{vi}(z, s) \perp\!\!\!\perp \tilde{Z}_v \mid \tilde{L}_v$ , where  $X_1 \perp\!\!\!\perp X_2 \mid X_3$  means that  $X_1$  and  $X_2$  are independent conditional on  $X_3$ .

Under Assumption 1, the inverse probability-weighted estimator for treatment  $z$  and strategy  $\alpha$  equals

$$\hat{Y}^{\text{ipw}}(z, \alpha) = n^{-1} \sum_{v=1}^m \sum_{i=1}^{N_v} \frac{Y_{vi} 1(Z_{vi} = z) \pi(S_{vi}; \alpha)}{f(\tilde{Z}_v \mid \tilde{L}_v)},$$

which can be expressed as a solution for  $\mu$  to the estimating equation  $\sum_{v=1}^m G_{z\alpha}^0(\tilde{Y}_v, \tilde{Z}_v, \tilde{L}_v; \mu) = 0$ , where

$$G_{z\alpha}^0(\tilde{Y}_v, \tilde{Z}_v, \tilde{L}_v; \mu) = \sum_{i=1}^{N_v} \left\{ \frac{Y_{vi} 1(Z_{vi} = z) \pi(S_{vi}; \alpha)}{f(\tilde{Z}_v \mid \tilde{L}_v)} - \mu \right\}.$$

Let  $\mu_{z\alpha}$  be the solution to  $\int G_{z\alpha}^0(\tilde{y}_v, \tilde{z}_v, \tilde{l}_v; \mu_{z\alpha}) dF(\tilde{y}_v, \tilde{z}_v, \tilde{l}_v) = 0$ . It is straightforward to show that  $\mu_{z\alpha} = k^{-1} E\{\sum_{i=1}^{N_v} \tilde{Y}_{vi}(z, \alpha)\}$ , where  $k = E(N_v)$  is the mean group size in the superpopulation and  $\tilde{Y}_{vi}(z, \alpha) = \sum_s Y_{vi}(z, s) \pi(s; \alpha)$ , with the summation being taken over all vectors  $s \in \{0, 1\}^{N_v-1}$ . If  $Y_v^*(z, \alpha) \perp\!\!\!\perp N_v$  where  $Y_v^*(z, \alpha) = \sum_{i=1}^{N_v} \tilde{Y}_{vi}(z, \alpha) / N_v$ , i.e., if the average potential outcome within a group is independent of the number of individuals within the group, then  $\mu_{z\alpha} = E\{Y_v^*(z, \alpha)\}$ . In other words,  $\mu_{z\alpha}$  is the mean group average potential outcome in the superpopulation, analogous to  $\bar{y}(z, \alpha)$  defined in § 2. Define the direct effect in the superpopulation by  $\bar{DE}(\alpha) = g(\mu_{1\alpha}, \mu_{0\alpha})$ ; the indirect, total and overall effects in the superpopulation can be defined analogously.

The Hájek-type estimators can also be expressed as solutions to estimation equations. Specifically, under Assumption 1,

$$\hat{Y}_h^{\text{haj}}(z, \alpha) = \hat{n}_{hz}^{-1} \sum_{v=1}^m \sum_{i=1}^{N_v} \frac{Y_{vi} 1(Z_{vi} = z) \pi(S_{vi}; \alpha)}{f(\tilde{Z}_v \mid \tilde{L}_v)} \quad (h = 1, 2),$$

where now

$$\hat{n}_{1z} = \sum_{v=1}^m \sum_{i=1}^{N_v} \frac{1(Z_{vi} = z)}{f(Z_{vi} \mid L_{vi})}, \quad \hat{n}_{2z} = \sum_{v=1}^m \sum_{i=1}^{N_v} \frac{1(Z_{vi} = z) \pi(S_{vi}; \alpha)}{f(\tilde{Z}_v \mid \tilde{L}_v)}.$$

It follows that  $\hat{Y}_h^{\text{haj}}(z, \alpha)$  solves  $\sum_{v=1}^m G_{z\alpha}^h(\tilde{Y}_v, \tilde{Z}_v, \tilde{L}_v; \mu) = 0$ , where

$$G_{z\alpha}^1(\tilde{Y}_v, \tilde{Z}_v, \tilde{L}_v; \mu) = \sum_{i=1}^{N_v} \left\{ \frac{Y_{vi} 1(Z_{vi} = z) \pi(S_{vi}; \alpha)}{f(\tilde{Z}_v | \tilde{L}_v)} - \mu \frac{1(Z_{vi} = z)}{f(Z_{vi} | L_{vi})} \right\},$$

$$G_{z\alpha}^2(\tilde{Y}_v, \tilde{Z}_v, \tilde{L}_v; \mu) = \sum_{i=1}^{N_v} \left\{ \frac{Y_{vi} 1(Z_{vi} = z) \pi(S_{vi}; \alpha)}{f(\tilde{Z}_v | \tilde{L}_v)} - \mu \frac{1(Z_{vi} = z) \pi(S_{vi}; \alpha)}{f(\tilde{Z}_v | \tilde{L}_v)} \right\}.$$

It is straightforward to show that  $\mu_{z\alpha}$  also satisfies  $\int G_{z\alpha}^h(\tilde{y}_v, \tilde{z}_v, \tilde{l}_v; \mu_{z\alpha}) dF(\tilde{y}_v, \tilde{z}_v, \tilde{l}_v) = 0$  ( $h = 1, 2$ ).

The asymptotic distributions of the inverse probability-weighted and Hájek-type estimators can be derived from standard estimating equation theory (Stefanski & Boos, 2002; Perez-Heydrich et al., 2014). For example, the proposition below establishes that the three direct effect estimators are asymptotically normal and gives closed-form expressions for the asymptotic variances when the propensity scores are known. The proposition entails the vector estimating equation  $G_{\alpha}^h(\tilde{Y}_v, \tilde{Z}_v, \tilde{L}_v; \theta) = \{G_{0\alpha}^h(\tilde{Y}_v, \tilde{Z}_v, \tilde{L}_v; \theta_1), G_{1\alpha}^h(\tilde{Y}_v, \tilde{Z}_v, \tilde{L}_v; \theta_2)\}^T$ , where  $\theta = (\theta_1, \theta_2)$ .

PROPOSITION 2. Suppose that Assumption 1 holds, the propensity scores are known, and the regularity assumptions in the Appendix hold. Then  $m^{1/2}\{\widehat{\text{DE}}^{\text{ipw}}(\alpha) - \bar{\text{DE}}(\alpha)\}$  converges in distribution to  $N(0, \Sigma_0^D)$  and  $m^{1/2}\{\widehat{\text{DE}}_h^{\text{haj}}(\alpha) - \bar{\text{DE}}(\alpha)\}$  converges in distribution to  $N(0, \Sigma_h^D)$  ( $h = 1, 2$ ) as  $m \rightarrow \infty$ , where

$$\Sigma_h^D = \tau U_h^{-1} V_h (U_h^{-1})^T \tau^T$$

with  $\tau = \{\partial g(x_1, x_0)/\partial x_1, \partial g(x_1, x_0)/\partial x_0\}$ ,  $U_h = E\{\partial G_{\alpha}^h(\tilde{Y}_v, \tilde{Z}_v, \tilde{L}_v; \theta_0)/\partial \theta_0\}$  and  $V_h = E\{G_{\alpha}^h(\tilde{Y}_v, \tilde{Z}_v, \tilde{L}_v; \theta_0)^{\otimes 2}\}$  ( $h = 0, 1, 2$ ); here  $\theta_0 = (\mu_{0\alpha}, \mu_{1\alpha})$ .

A comparison between  $\Sigma_h^D$  ( $h = 1, 2$ ) and  $\Sigma_0^D$  explains why the Hájek-type estimators can vary less than the inverse probability-weighted estimator. For example, suppose that the contrast  $g$  is the difference function. Denote  $G_{z\alpha}^0(\tilde{Y}_v, \tilde{Z}_v, \tilde{L}_v; \theta_0)$  by  $G_{z\alpha}^0$  and note that  $\Sigma_0^D = E(G_{1\alpha}^0 - G_{0\alpha}^0)^2/k^2$  and  $\Sigma_h^D = E(G_{1\alpha}^0 - G_{0\alpha}^0 - W_h)^2/k^2$  ( $h = 1, 2$ ), where  $W_h = \mu_{1\alpha}\{\hat{N}_{hv}(1, \alpha) - N_v\} - \mu_{0\alpha}\{\hat{N}_{hv}(0, \alpha) - N_v\}$  with  $\hat{N}_{1v}(z, \alpha) = \sum_{i=1}^{N_v} 1(Z_{vi} = z)/f(Z_{vi} | L_{vi})$  and  $\hat{N}_{2v}(z, \alpha) = \sum_{i=1}^{N_v} 1(Z_{vi} = z)\pi(S_{vi}; \alpha)/f(\tilde{Z}_v | \tilde{L}_v)$ . Thus, the Hájek estimators will have smaller asymptotic variance if and only if  $\text{var}(W_h) < 2E\{(G_{1\alpha}^0 - G_{0\alpha}^0)W_h\}$ , and so are expected to be less variable when  $G_{1\alpha}^0 - G_{0\alpha}^0$  and  $W_h$  are strongly correlated. In the extreme scenario of  $Y_{vi}(z, s) = c_z$  ( $v = 1, \dots, m; i = 1, \dots, N_v$ ), we have  $W_2 = G_{1\alpha}^0 - G_{0\alpha}^0$  and  $\Sigma_2^D = 0$  but  $\Sigma_0^D > 0$  in general.

In observational studies, the mechanism by which individuals select treatment is in general not known, so that  $f(\tilde{Z}_v | \tilde{L}_v)$  and  $f(Z_{vi} | L_{vi})$  must be estimated in order to construct inverse probability-weighted estimators. In practice, due to the curse of dimensionality, one might assume a parametric model for the propensity scores (Tchetgen Tchetgen & VanderWeele, 2012). Let  $G(\tilde{Z}_v, \tilde{L}_v; \gamma)$  denote the score function for the likelihood under the assumed propensity score model indexed by a finite-dimensional parameter vector  $\gamma$ , and let  $\gamma_0$  denote the true parameter value, which is the solution to  $\int G(\tilde{z}_v, \tilde{l}_v; \gamma) dF(\tilde{z}_v, \tilde{l}_v; \gamma) = 0$ . Now consider the vector estimating equation  $G_{\alpha}^{*h}(\tilde{Y}_v, \tilde{Z}_v, \tilde{L}_v; \theta) = \{G_{0\alpha}^h(\tilde{Y}_v, \tilde{Z}_v, \tilde{L}_v; \theta_1), G_{1\alpha}^h(\tilde{Y}_v, \tilde{Z}_v, \tilde{L}_v; \theta_2), G(\tilde{Z}_v, \tilde{L}_v; \theta_3)\}^T$  ( $h = 0, 1, 2$ ) where  $\theta = (\theta_1, \theta_2, \theta_3)$ .

PROPOSITION 3. *Suppose that Assumption 1 holds, the parametric propensity score model is correctly specified, and the regularity assumptions in the Appendix hold. Then  $m^{1/2}\{\widehat{DE}^{ipw}(\alpha) - \bar{DE}(\alpha)\}$  converges in distribution to  $N(0, \Sigma_0^{*D})$  and  $m^{1/2}\{\widehat{DE}_h^{haj}(\alpha) - \bar{DE}(\alpha)\}$  converges in distribution to  $N(0, \Sigma_h^{*D})$  ( $h = 1, 2$ ) as  $m \rightarrow \infty$ , where*

$$\Sigma_h^{*D} = \tau^* U_h^{*-1} V_h^* (U_h^{-1})^T \tau^{*T}$$

with  $\tau^* = (\tau, 0_p)$ ,  $U_h^* = E\{\partial G_\alpha^{*h}(\tilde{Y}_v, \tilde{Z}_v, \tilde{L}_v; \theta_0)/\partial \theta_0\}$  and  $V_h^* = E\{G_\alpha^{*h}(\tilde{Y}_v, \tilde{Z}_v, \tilde{L}_v; \theta_0)^{\otimes 2}\}$  ( $h = 0, 1, 2$ ); here  $\theta_0 = (\mu_{0\alpha}, \mu_{1\alpha}, \gamma_0)$ ,  $0_p$  denotes the  $1 \times p$  zero vector, and  $p$  is the dimension of  $\gamma$ .

Proposition 3 establishes the asymptotic normality of  $\widehat{DE}^{ipw}(\alpha)$ ,  $\widehat{DE}_1^{haj}(\alpha)$  and  $\widehat{DE}_2^{haj}(\alpha)$  when the propensity score is correctly modelled. The asymptotic variance can be estimated consistently using empirical sandwich estimators, i.e., by replacing  $U_h^*$  and  $V_h^*$  with their empirical counterparts (Stefanski & Boos, 2002). In the Appendix the asymptotic variance of  $\widehat{DE}^{ipw}(\alpha)$  when the propensity score is estimated is shown to be no greater than when the propensity score is known. This is analogous to the well-known result about weighted estimators in the absence of interference; that is, even if the propensity scores are known, it is more efficient to use estimates of the propensity scores when computing inverse probability-weighted estimators. This relationship between the asymptotic variances when the propensity scores are known and when they are unknown but correctly modelled also holds for the Hájek-type estimators. Asymptotic normality of the indirect, total and overall effect estimators can be derived similarly.

## 5. SIMULATION STUDY

A simulation study was conducted to investigate the bias, empirical standard error and average estimated standard error of the different estimators discussed in § 4. In the simulations the inverse probability-weighted and Hájek-type effect estimators were computed using the true propensity score, an estimated propensity score based on a correct model, and an estimated propensity score based on a misspecified model. Simulations were conducted under partial interference, i.e., Assumption 1, for both continuous and binary outcomes. The simulation study for a continuous outcome was carried out in the steps described below.

*Step 1.* A random sample of  $m = 500$  groups was created as follows. First, the group size  $N_v$  was randomly sampled from  $\{2, 3, 4, 5, 6\}$  with corresponding probabilities  $1/8, 1/8, 1/2, 3/16, 1/16$ . For each individual in each group,  $\varepsilon_{vi}$  was randomly sampled from  $N(0, 1)$  ( $v = 1, \dots, m$ ;  $i = 1, \dots, N_v$ ). Then the potential outcomes for individual  $i$  in group  $v$  were set to  $Y_{vi}(z_{vi}, s_{vi}) = 5 + 3z_{vi} + 2 \sum s_{vi} + \varepsilon_{vi}$ .

*Step 2.* The covariate vectors  $L_{vi} = (L_{vi1}, \dots, L_{vi4})$  were randomly sampled from  $N(0, I_4)$  ( $v = 1, \dots, m$ ;  $i = 1, \dots, N_v$ ), where  $I_4$  denotes the  $4 \times 4$  identity matrix.

*Step 3.* Treatment variables  $Z_{vi}$  were simulated from a Bernoulli distribution with mean  $\text{logit}^{-1}(\gamma_0 + \gamma_1 L_{vi1} + \gamma_2 L_{vi2} + \gamma_3 L_{vi3} + \gamma_4 L_{vi4} + b_v)$ , where the random effects  $b_v$  were randomly sampled from  $N(0, 1)$  ( $v = 1, \dots, m$ ) and  $(\gamma_0, \gamma_1, \gamma_2, \gamma_3, \gamma_4) = (0.5, -1, 0.5, -0.25, -0.1)$ .

*Step 4.* A correctly specified logistic regression model  $\text{logit}\{f(1 | b, L)\} = \gamma_0 + \gamma_1 L_{vi1} + \gamma_2 L_{vi2} + \gamma_3 L_{vi3} + \gamma_4 L_{vi4} + b_v$  and a misspecified logistic regression model  $\text{logit}\{f(1 | b, L)\} = \gamma_0 +$



Table 1. Empirical bias ( $\times 10$ ), empirical standard error, and average estimated standard error of the estimators of  $\bar{D}E(\alpha)$  with a continuous outcome

Known $f$	$\alpha = 0.1$			$\alpha = 0.5$			$\alpha = 0.9$		
	Bias	ESE	ASE	Bias	ESE	ASE	Bias	ESE	ASE
$\hat{D}E^{ipw}(\alpha)$	0.4	1.4	1.4	0.1	0.7	0.7	0.2	1.7	1.7
$\hat{D}E_1^{haj}(\alpha)$	0.6	1.5	1.5	0.1	0.6	0.6	0.5	1.6	1.6
$\hat{D}E_2^{haj}(\alpha)$	0.1	0.3	0.3	0.0	0.2	0.2	0.0	0.3	0.3
Correct $f$	Bias	ESE	ASE	Bias	ESE	ASE	Bias	ESE	ASE
$\hat{D}E^{ipw}(\alpha)$	4.1	1.5	1.4	0.7	0.6	0.6	7.3	1.3	1.3
$\hat{D}E_1^{haj}(\alpha)$	3.8	1.5	1.5	0.5	0.6	0.8	7.6	1.3	1.5
$\hat{D}E_2^{haj}(\alpha)$	0.2	0.3	0.3	0.8	0.2	0.2	0.5	0.3	0.3
Mis $f$	Bias	ESE	ASE	Bias	ESE	ASE	Bias	ESE	ASE
$\hat{D}E^{ipw}(\alpha)$	0.4	1e1	1e1	20	1e3	1e3	10	2e2	1e2
$\hat{D}E_1^{haj}(\alpha)$	5.2	2e0	1e1	10	1e3	1e3	30	3e2	2e2
$\hat{D}E_2^{haj}(\alpha)$	0.3	0.3	0.3	0.8	0.3	0.3	0.3	0.5	0.5

ESE, empirical standard error; ASE, average estimated standard error; Known  $f$ , true propensity score known; Correct  $f$ , propensity score unknown but correctly modelled; Mis  $f$ , propensity score incorrectly modelled.

$\gamma_1 X_{vi1} + \gamma_2 X_{vi2} + \gamma_3 X_{vi3} + \gamma_4 X_{vi4} + b_v$ , where  $X_{vi1} = \exp(L_{vi1}/2)$ ,  $X_{vi2} = L_{vi2}/\{1 + \exp(L_{vi1})\} + 1$ ,  $X_{vi3} = (L_{vi1}L_{vi3}/1.5 + 0.6)^3$  and  $X_{vi4} = (L_{vi1} + L_{vi4} + 2)^2$ , were fitted to the simulated data.

*Step 5.* The causal effect estimators and their corresponding variance estimators were calculated for  $\alpha_1 = 0.1, 0.5, 0.9$  and  $\alpha_0 = 0.1$  using the known propensity score, the estimated propensity score from the correctly specified mixed-effects model and the estimated propensity score from the misspecified mixed-effects model.

*Step 6.* Steps 1–5 were repeated 10 000 times, and the empirical bias, empirical standard error and average estimated standard error were calculated for the estimators in Step 5.

From the potential outcome model specified in Step 1 it follows that  $\mu_{z\alpha} = 5 + 3z + 2(\eta - 1)\alpha$  and  $\mu_\alpha = 5 + (2\eta + 1)\alpha$  where  $\eta = E(N_v^2)/E(N_v)$ . Hence  $\bar{D}E(\alpha) = 3$  for any  $\alpha \in (0, 1)$ ,  $\bar{I}E(\alpha_1, \alpha_0) = 2(\eta - 1)(\alpha_1 - \alpha_0)$ ,  $\bar{T}E(\alpha_1, \alpha_0) = 3 + 2(\eta - 1)(\alpha_1 - \alpha_0)$  and  $\bar{O}E(\alpha_1, \alpha_0) = (2\eta + 1)(\alpha_1 - \alpha_0)$ . Simulation results for the direct effect estimators are given in Table 1. All three estimators are approximately unbiased when the propensity scores are known or correctly modelled, but are biased if the propensity scores are incorrectly modelled. For all three estimators the average estimated standard error is also relatively close to the empirical standard error when the propensity scores are known or correctly modelled. Note that  $\hat{D}E_2^{haj}(\alpha)$  has substantially smaller empirical standard error than  $\hat{D}E^{ipw}(\alpha)$  and  $\hat{D}E_1^{haj}(\alpha)$ . For example, when  $\alpha = 0.1$  and the propensity scores are known, the empirical standard errors of  $\hat{D}E^{ipw}(\alpha)$  and  $\hat{D}E_1^{haj}(\alpha)$  are 1.4 and 1.5, whereas the empirical standard error of  $\hat{D}E_2^{haj}(\alpha)$  is only 0.3. Similar results hold when the propensity scores are treated as unknown and either correctly or incorrectly modelled. The results in Table 1 demonstrate that, as well as having smaller empirical standard error,  $\hat{D}E_2^{haj}(\alpha)$  may be more robust than  $\hat{D}E^{ipw}(\alpha)$  and  $\hat{D}E_1^{haj}(\alpha)$  with respect to misspecification of the propensity score model.

Table 2. Empirical bias, empirical standard error, and average estimated standard error of the estimators of  $\widehat{DE}(\alpha)$  with a binary outcome; all values have been multiplied by 100

	$\alpha = 0.1$			$\alpha = 0.5$			$\alpha = 0.9$		
Known $f$	Bias	ESE	ASE	Bias	ESE	ASE	Bias	ESE	ASE
$\widehat{DE}^{ipw}(\alpha)$	0.2	9.7	9.7	0	4.7	4.8	0.1	9.3	9.2
$\widehat{DE}_1^{haj}(\alpha)$	0.3	9.6	9.7	0	4.6	4.5	0.1	8.4	8.4
$\widehat{DE}_2^{haj}(\alpha)$	0.1	7.0	6.9	0	3.9	3.9	0	5.4	5.3
Correct $f$	Bias	ESE	ASE	Bias	ESE	ASE	Bias	ESE	ASE
$\widehat{DE}^{ipw}(\alpha)$	1.4	10.2	9.8	0.1	4.5	4.3	3.5	7.4	7.2
$\widehat{DE}_1^{haj}(\alpha)$	1.3	10.1	9.8	0.2	4.5	4.5	3.5	7.3	7.5
$\widehat{DE}_2^{haj}(\alpha)$	0.1	7.3	7.1	0.5	3.8	3.6	0.9	5.2	5.0
Mis $f$	Bias	ESE	ASE	Bias	ESE	ASE	Bias	ESE	ASE
$\widehat{DE}^{ipw}(\alpha)$	1	1e2	1e2	1	5e2	3e2	1e1	3e4	2e4
$\widehat{DE}_1^{haj}(\alpha)$	1	4e1	1e2	10	5e2	3e2	2e1	3e4	2e4
$\widehat{DE}_2^{haj}(\alpha)$	0.1	7.4	7.3	1.2	6.8	6.4	1	9.7	9.5

The simulation study described above was repeated for a binary outcome. Specifically, Step 1 was replaced with the following, while all other steps remained the same.

*Step 1.* A random sample of  $m = 500$  groups was created as follows. First, the group size  $N_v$  was randomly sampled from  $\{2, 3, 4, 5\}$  with corresponding probabilities  $1/8, 1/8, 1/2, 1/4$ . Then the potential outcomes  $Y_{vi}(z_{vi}, s_{vi})$  were set to 0 with probability 0.2, 1 with probability 0.2, and  $1(Z_{vi} = 1, \sum S_{vi} = |s_{vi}|)$  ( $v = 1, \dots, m; i = 1, \dots, N_v$ ) with probability 0.6.

For this potential outcome model,  $\mu_{1\alpha} = 0.6\lambda + 0.2$ ,  $\mu_{0\alpha} = 0.2$  and  $\mu_\alpha = 0.6\alpha\lambda + 0.2$  with  $\lambda = E(\alpha^{N_v-1}N_v)/E(N_v)$ . Simulation results for this scenario are given in Table 2. Similar to the continuous outcome simulations, the empirical standard error for  $\widehat{DE}_2^{haj}(\alpha)$  is smaller than that for  $\widehat{DE}^{ipw}(\alpha)$  and  $\widehat{DE}_1^{haj}(\alpha)$  in all three scenarios, and  $\widehat{DE}_2^{haj}(\alpha)$  also tends to be more robust with respect to misspecification of the propensity score model than the other two estimators. Similar results, not shown here, were observed for the other causal effect estimators.

## 6. ROTAVIRUS VACCINE STUDY IN NICARAGUA

Rotavirus diarrhoea is a major health problem in Nicaragua (Espinoza et al., 1997). The pentavalent rotavirus vaccine was introduced in 2006. Nicaraguan infants are offered the vaccine at two, four and six months of age as part of the country's Expanded Program on Immunization. In 2010, a study to assess the impact of the immunization programme was carried out in León, Nicaragua's second largest city, with an estimated population in 2010 of close to 200 000. The Health and Demographic Surveillance Site-León was employed to obtain a simple random sample of households from 50 out of 208 randomly selected geographical clusters of equal size in León (Becker-Dreps et al., 2013). For simplicity, in the following analysis the cluster sampling used to obtain these data is ignored. There were 530 households in the study, and any child in a selected household under the age of five was eligible to participate. Information was collected about each household, including water source, sanitation system, maternal education level, and the dates of

Table 3. Effect estimates for the rotavirus vaccine study; all values have been multiplied by 10

	$\alpha = 0.2$		$\alpha = 0.4$		$\alpha = 0.6$		$\alpha = 0.8$	
	Est	SE	Est	SE	Est	SE	Est	SE
$\hat{DE}^{ipw}(\alpha)$	-0.79	1.30	-0.62	0.97	-0.42	0.67	-0.18	0.44
$\hat{DE}_1^{haj}(\alpha)$	-0.81	1.31	-0.63	0.98	-0.43	0.68	-0.20	0.45
$\hat{DE}_2^{haj}(\alpha)$	-0.52	1.11	-0.44	0.84	-0.32	0.59	-0.14	0.41
$\hat{IE}^{ipw}(\alpha, 0.1)$	-0.13	0.17	-0.40	0.50	-0.69	0.84	-0.98	1.18
$\hat{IE}_1^{haj}(\alpha, 0.1)$	-0.13	0.17	-0.41	0.50	-0.69	0.84	-0.99	1.18
$\hat{IE}_2^{haj}(\alpha, 0.1)$	-0.04	0.16	-0.15	0.45	-0.28	0.73	-0.45	1.01
$\hat{TE}^{ipw}(\alpha, 0.1)$	-0.93	1.46	-1.02	1.45	-1.11	1.45	-1.17	1.45
$\hat{TE}_1^{haj}(\alpha, 0.1)$	-0.94	1.47	-1.04	1.46	-1.12	1.46	-1.18	1.46
$\hat{TE}_2^{haj}(\alpha, 0.1)$	-0.56	1.25	-0.59	1.24	-0.61	1.25	-0.59	1.25
$\hat{OE}^{ipw}(\alpha, 0.1)$	-0.20	0.28	-0.56	0.73	-0.85	1.05	-1.04	1.25
$\hat{OE}_1^{haj}(\alpha, 0.1)$	-0.20	0.28	-0.57	0.73	-0.86	1.06	-1.05	1.25
$\hat{OE}_2^{haj}(\alpha, 0.1)$	-0.09	0.24	-0.27	0.63	-0.42	0.91	-0.51	1.08

Est, point estimate; SE, estimated standard error.

birth of study participants. Each individual in the study was visited fortnightly by a fieldworker for approximately one year. At each visit information about diarrhoea episodes in the past 14 days was recorded. The primary outcome  $Y$  was whether a child had at least one diarrhoea episode during the study.

For each child we assumed their interference set to be other children in the same household. A mixed-effects logistic regression model of the probability of having received all three scheduled doses was fitted conditional on the following baseline covariates: child's age, categorized as 0–11 months, 12–23 months, or 24–59 months; mother's education level, categorized as primary education only or at least some secondary education; dirt household floor or not; dry or wet season; household indoor toilet, latrine, or none; indoor municipal water supply or not; and breastfeeding or not. Likelihood ratio tests from the fitted logistic model indicated that the odds of having all three doses of vaccine was higher among children whose mothers were more educated, with  $p = 0.01$ .

Effect estimates and estimated standard errors are reported in Table 3 for the inverse probability-weighted and the two Hájek estimators for contrast function  $g(x_1, x_0) = x_1 - x_0$ . The Hájek 2 estimates are closer to the null value of zero and, as expected, have 15–20% smaller estimated standard errors than the inverse probability-weighted and Hájek 1 estimates. The direct effect estimates indicate the expected difference in the proportions of children who will acquire rotavirus diarrhoea among vaccinated versus unvaccinated children for a fixed level of vaccine coverage  $\alpha$ . The estimated direct effects become closer to the null as  $\alpha$  increases, suggesting that the direct protective effect of vaccination decreases as additional children in the household are vaccinated. The indirect effect estimates approximate the expected difference in the proportions of unvaccinated children who will acquire diarrhoea when vaccine coverage is  $\alpha\%$  versus 10%. The total effect estimates indicate the expected difference in the proportions of vaccinated children who will acquire diarrhoea when vaccine coverage is  $\alpha\%$  compared with unvaccinated children when vaccine coverage is 10%. The overall effect estimates provide simple summary comparisons between any two allocation strategies; for example, according to the Hájek 2 estimates, 5.1 fewer cases of diarrhoea per 100 individuals per year would be expected if on average 80% of children in a household were vaccinated than if on average only 10% of children were vaccinated.

7. DISCUSSION

The inverse probability-weighted estimator and two Hájek-type estimators in § 3 allow for any form of interference between individuals, with the former being unbiased in a finite-population model with known propensity scores. Assuming partial interference and random sampling of groups from a superpopulation, all three estimators are consistent and asymptotically normal when the propensity scores are known or correctly modelled. Empirical results demonstrate that the second Hájek estimator can have substantially smaller finite-sample variance than the other two estimators. One avenue of future research entails deriving the estimators’ large-sample properties without assuming partial interference. Another future direction might involve developing estimators which are robust with respect to misspecification of the propensity score model. Throughout this work conditional exchangeability is assumed, i.e., treatment is assumed to be independent of potential outcomes conditional on an observable set of covariates. In future work one could investigate relaxing this assumption, perhaps via sensitivity analysis or instrumental variable methods. Finally, the target parameters in this paper utilize the Bernoulli allocation strategy proposed by Tchetgen Tchetgen & VanderWeele (2012). These estimands consider the counterfactual scenario where individuals independently select treatment with equal probability. In scenarios where interference is present, it is unlikely that individual treatment selections would be independent. Therefore further interference-related research might target alternative parameters.

ACKNOWLEDGEMENT

The authors were partially supported by the U.S. National Institutes of Health. The fieldwork was supported by the Thrasher Research Fund. The content of this paper is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The authors thank M. Elizabeth Halloran, Joseph Rigdon, an associate editor, and a reviewer for helpful comments.

APPENDIX

*Proof of Proposition 1*

To show that  $\hat{Y}^{ipw}(z, \alpha)$  is unbiased, observe that

$$\begin{aligned} E\{\hat{Y}^{ipw}(z, \alpha)\} &= n^{-1} \sum_i \sum_{z_i, s_i} \frac{y_i(z_i, s_i) 1(z_i = z) \pi(s_i; \alpha)}{f(z_i, s_i | l_i, l_{x_i})} f(z_i, s_i | l_i, l_{x_i}) \\ &= n^{-1} \sum_i \sum_{s_i} y_i(z, s_i) \pi(s_i; \alpha) = \bar{y}(z, \alpha). \end{aligned}$$

That  $\hat{Y}^{ipw}(\alpha)$  is unbiased can be proved similarly.

*Proof of Propositions 2 and 3*

To prove Proposition 2, assume that there exist constants  $c_1, c_2, c_3 < \infty$  and  $\delta > 0$  such that  $-c_1 < Y_{vi} < c_2, N_v < c_3, \delta < f(\tilde{Z}_v | \tilde{L}_v)$  and  $\delta < f(Z_{vi} | L_{vi})$  with probability 1. Let  $\hat{\theta}^0 = \{\hat{Y}^{ipw}(0, \alpha), \hat{Y}^{ipw}(1, \alpha)\}$  and  $\hat{\theta}^h = \{\hat{Y}_h^{haj}(0, \alpha), \hat{Y}_h^{haj}(1, \alpha)\}$  ( $h = 1, 2$ ). Let  $G^h(\theta) = \{G_0^h(\theta), G_1^h(\theta)\}^T$  denote the vector estimating equation  $G_\alpha^h(\tilde{Y}_v, \tilde{Z}_v, \tilde{L}_v; \theta)$ . Let  $\dot{G}^h(\theta_0) = \partial G^h(\theta_0) / \partial \theta_0^T$  and write  $\|v\|^2 = v_1^2 + \dots + v_p^2$  for any vector  $v$  of length  $p$ .

First we show that the following four conditions hold for  $h = 0, 1, 2$ : (i)  $E\{\dot{G}^h(\theta_0)\}$  exists and is nonsingular; (ii)  $G^h(\theta)$  is twice continuously differentiable with respect to  $\theta$  for every  $(\tilde{y}_v, \tilde{z}_v, \tilde{l}_v)$ ;

(iii)  $|\partial^2 G_z^h(\theta)/(\partial\theta_i\partial\theta_j)| \leq \psi$  for some integrable measurable function  $\psi$ ; and (iv)  $E\|G^h(\theta_0)\|^2 < \infty$ . It is straightforward to show that  $E\{\dot{G}^h(\theta_0)\} = -I_2 E(N_v)$ , where  $I_2$  is the  $2 \times 2$  identity matrix, implying (i). Note that  $\partial^2 G_z^h(\theta)/(\partial\theta_i\partial\theta_j) = 0$ , so (ii) holds and (iii) is satisfied for the function  $\psi = 0$ . To show (iv), observe that

$$E\|G^0(\theta_0)\|^2 = \sum_{z=0}^1 E \left\{ \sum_{i=1}^{N_v} \frac{Y_{vi} 1(Z_{vi} = z) \pi(S_{vi}; \alpha)}{f(\tilde{Z}_v | \tilde{L}_v)} - N_v \mu_{z\alpha} \right\}^2.$$

From the boundedness assumptions on  $Y_{vi}$ ,  $N_v$  and  $f(\tilde{Z}_v | \tilde{L}_v)$ , it follows that  $E\|G^0(\theta_0)\|^2 < \infty$ . Similar results can be established for  $h = 1, 2$ .

Next, note that  $G_{z\alpha}^0(\tilde{Y}_v, \tilde{Z}_v, \tilde{L}_v; \mu_{z\alpha})$  is a linear function of  $\mu_{z\alpha}$  with slope  $-N_v$ . For  $h = 1, 2$ ,  $G_{z\alpha}^h(\tilde{Y}_v, \tilde{Z}_v, \tilde{L}_v; \mu_{z\alpha})$  is also a linear function of  $\mu_{z\alpha}$  with finite, nonzero slope, because by assumption  $f(\tilde{Z}_v | \tilde{L}_v) > 0$  and there exists at least one  $i$  such that  $Z_{vi} = z$ . Hence, the solution for  $\theta$  to  $\sum_{v=1}^m G^h(\theta) = 0$  is unique for  $h = 0, 1, 2$ . Therefore, because (i)–(iv) hold, by Theorem 5.4.2 of van der Vaart (1998),  $\hat{\theta}^h$  converges in probability to  $\theta_0$ . Proposition 2 then follows from Theorem 5.4.1 of van der Vaart (1998) and the delta method.

Similar reasoning can be used to prove Proposition 3 under the following additional assumptions about the parametric propensity score model:  $\gamma_0$  is in an open subset of Euclidean space;  $E\{\dot{G}(\gamma_0)\}$  exists and is nonsingular, where  $\dot{G}(\gamma_0) = \partial G(\tilde{Z}_v, \tilde{L}_v; \gamma_0)/\partial \gamma_0^T$ ;  $G(\tilde{z}_v, \tilde{l}_v; \gamma)$  is twice continuously differentiable with respect to  $\gamma$  and  $|\partial^2 G(\tilde{z}_v, \tilde{l}_v; \gamma)/(\partial \gamma_i \partial \gamma_j)| \leq \psi$  for some integrable measurable function  $\psi$  for every  $(\tilde{z}_v, \tilde{l}_v)$ ; and  $E\|G(\tilde{Z}_v, \tilde{L}_v; \gamma_0)\|^2 < \infty$ .

*Proof of reduction in variance with a correctly specified propensity score model*

Using block matrix notation, write

$$U_0^* = \begin{pmatrix} U_0 & U_{\mu\gamma} \\ 0_{p \times 2} & U_\gamma \end{pmatrix}, \quad V_0^* = \begin{pmatrix} V_0 & V_{\mu\gamma} \\ V_{\mu\gamma}^T & V_\gamma \end{pmatrix},$$

where  $0_{p \times 2}$  is the  $p \times 2$  matrix of zeros. It is straightforward to show that  $U_\gamma = -V_\gamma$  and  $U_{\mu\gamma} = -V_{\mu\gamma}$ . It follows that

$$U_0^{*-1} = \begin{pmatrix} U_0^{-1} & -U_0^{-1} V_{\mu\gamma} V_\gamma^{-1} \\ 0_{p \times 2} & U_\gamma^{-1} \end{pmatrix}$$

and therefore

$$U_0^{*-1} V_0^* (U_0^{*-1})^T = \left\{ \begin{array}{cc} U_0^{-1} V_0 (U_0^{-1})^T - U_0^{-1} V_{\mu\gamma} V_\gamma^{-1} V_{\mu\gamma}^T (U_0^{-1})^T & \star \\ \star & \star \end{array} \right\},$$

where  $\star$  denotes quantities not expressed explicitly. Hence

$$\Sigma_0^{*D} = \Sigma_0^D - \tau U_0^{-1} V_{\mu\gamma} V_\gamma^{-1} V_{\mu\gamma}^T (U_0^{-1})^T \tau^T.$$

Since  $V_\gamma = E\{G(\tilde{Z}_v, \tilde{L}_v; \gamma_0)^{\otimes 2}\}$  is positive semidefinite, so is  $V_\gamma^{-1}$ . Therefore  $\Sigma_0^D \geq \Sigma_0^{*D}$ . The same approach can be used to show that  $\Sigma_h^D \geq \Sigma_h^{*D}$  for  $h = 1, 2$ .

REFERENCES

BECKER-DREPS, S., MELÉNDEZ, M., LIU, L., ZAMBRANA, L. E., PANIAGUA, M., WEBER, D. J., HUDGENS, M. G., CÁCERES, M., KÄLLESTÄLL, C., MORGAN, D. R. et al. (2013). Community diarrhea incidence before and after rotavirus vaccine introduction in Nicaragua. *Am. J. Trop. Med. Hygiene* **89**, 246–50.

- ESPIÑOZA, F., PANIAGUA, M., HALLANDER, H., SVENSSON, L. & STRANNEGÅRD, Ö. (1997). Rotavirus infections in young Nicaraguan children. *Pediatric Inf. Dis. J.* **16**, 564–71.
- HÁJEK, J. (1971). Comment on a paper by D. Basu. In *Foundations of Statistical Inference*, V. Godambe & D. Sprott, eds. Toronto: Holt, Rinehart and Winston, p. 236.
- HALLORAN, M. E. & HUDGENS, M. G. (2012). Causal inference for vaccine effects on infectiousness. *Int. J. Biostatist.* **8**, 1–40.
- HALLORAN, M. E. & STRUCHINER, C. J. (1995). Causal inference in infectious diseases. *Epidemiology* **6**, 142–51.
- HONG, G. & RAUDENBUSH, S. W. (2006). Evaluating kindergarten retention policy: A case study of causal inference for multilevel observational data. *J. Am. Statist. Assoc.* **101**, 901–10.
- HORVITZ, D. G. & THOMPSON, D. J. (1952). A generalization of sampling without replacement from a finite universe. *J. Am. Statist. Assoc.* **47**, 663–85.
- HUDGENS, M. G. & HALLORAN, M. E. (2008). Toward causal inference with interference. *J. Am. Statist. Assoc.* **103**, 832–42.
- LIU, L. & HUDGENS, M. G. (2014). Large sample randomization inference of causal effects in the presence of interference. *J. Am. Statist. Assoc.* **109**, 288–301.
- LUO, X., SMALL, D. S., LI, C. S. R. & ROSENBAUM, P. R. (2012). Inference with interference between units in an fMRI experiment of motor inhibition. *J. Am. Statist. Assoc.* **107**, 530–41.
- MANSKI, C. F. (2013). Identification of treatment response with social interactions. *Economet. J.* **16**, S1–23.
- NICKERSON, D. W. (2008). Is voting contagious? Evidence from two field experiments. *Am. Polit. Sci. Rev.* **102**, 49–57.
- PÉREZ-HEYDRICH, C., HUDGENS, M. G., HALLORAN, M. E., CLEMENS, J. D., ALI, M. & EMCH, M. E. (2014). Assessing effects of cholera vaccination in the presence of interference. *Biometrics* **70**, 731–44.
- ROSENBAUM, P. R. (2007). Interference between units in randomized experiments. *J. Am. Statist. Assoc.* **102**, 191–200.
- SOBEL, M. E. (2006). What do randomized studies of housing mobility demonstrate? Causal inference in the face of interference. *J. Am. Statist. Assoc.* **101**, 1398–407.
- STEFANSKI, L. A. & BOOS, D. D. (2002). The calculus of M-estimation. *Am. Statistician* **56**, 29–38.
- TCHETGEN TCHETGEN, E. J. & VANDERWEELE, T. J. (2012). On causal inference in the presence of interference. *Statist. Meth. Med. Res.* **21**, 55–75.
- van der VAART, A. (1998). *Asymptotic Statistics*. Cambridge: Cambridge University Press.
- VANDERWEELE, T. J. & TCHETGEN TCHETGEN, E. J. (2011). Effect partitioning under interference in two-stage randomized vaccine trials. *Statist. Prob. Lett.* **81**, 861–9.
- VANDERWEELE, T. J., TCHETGEN TCHETGEN, E. J. & HALLORAN, M. E. (2012). Components of the indirect effect in vaccine trials: Identification of contagion and infectiousness effects. *Epidemiology* **23**, 751–61.

[Received December 2015. Revised August 2016]