Published in final edited form as:

Stat Probab Lett. 2011 July 1; 81(7): 861–869. doi:10.1016/j.spl.2011.02.019.

Effect partitioning under interference in two-stage randomized vaccine trials

Tyler J. VanderWeele and Eric J. Tchetgen Tchetgen¹

Departments of Epidemiology and Biostatistics, Harvard School of Public Health

Abstract

In the presence of interference, the exposure of one individual may affect the outcomes of others. We provide new effect partitioning results under interferences that express the overall effect as a sum of (i) the indirect (or spillover) effect and (ii) a contrast between two direct effects.

Keywords

Causal inference; counterfactual; interference; SUTVA; randomized experiments; spillover effects; vaccine trials

1. Introduction

In vaccine trials, the effect of a vaccine intervention on an individual's disease outcome depends not only on whether an individual receives a vaccine but may also depend on what proportion of the population in the individual's community receives the vaccine. In such settings when the outcome of one individual depends on the treatment or exposure status of other individuals, interference is said to be present. Drawing inferences about cause and effect is considerably more challenging in settings with such interference (Hong and Raudenbush, 2006; Sobel, 2006; Rosenbaum, 2007; Hudgens and Halloran, 2008; VanderWeele, 2010; Tchetgen Tchetgen and VanderWeele, 2011). One design that facilitates causal inference in the presence of interference in the vaccine context is a two-stage randomized experiment in which specific clusters are randomized to having a certain proportion of the cluster vaccinated (e.g. 50% versus 30%) and then, within each cluster, once the proportion is determined, individuals are randomized to receive the vaccine or not. In the second stage of such randomized experiments, either a fixed proportion of individuals are vaccinated (what we will call type A) or alternatively each individual is independently randomized to have the vaccine with some fixed probability (what we will call type B). Such designs are sometimes also referred to as split-plot or pseudo-cluster randomized designs.

In the presence of such interference, for each possible allocation of vaccine, it is of interest to know the extent to which an individual is protected as a result of receiving a vaccine and the extent to which the individual is protected due to others in the same cluster receiving the vaccine. The distinction between these two is sometimes referred to as the direct and indirect

© 2011 Elsevier B.V. All rights reserved.

¹Correspondence: Department of Epidemiology, Harvard School of Public Health, 677 Huntington, Boston MA 02115, tvanderw@hsph.harvard.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

effects of a vaccine respectively. Such issues arise, for example, in attempting to assess the extent of herd immunity conferred by a cholera vaccine (Ali et al., 2005) and we consider an example similar to this as an illustration later in the paper. Hudgens and Halloran (2008) showed how under type A randomization one could define and identify effects they called the overall, total, direct and indirect effects even in the presence of interference; it is also possible to obtain finite sample confidence intervals for these effects under type A randomization (Tchetgen Tchetgen and VanderWeele, 2011).

Building on the work of Halloran and Struchiner (1995), Hudgens and Halloran (2008) also formally showed that the total effect could be partitioned into the sum of what they defined as a direct effect and an indirect effect. Suppose in the first stage of the randomized experiment, clusters are randomized to receive either allocation program α_0 or α_1 . The total effect, as defined by Hudgens and Halloran (2008), is the effect on an individual comparing two scenarios, one in which the individual receives the vaccine and is in a cluster that receives allocation α_1 and the second in which the individual does not receive the vaccine and is in a cluster that receives allocation α_0 ; the population average total effect is then simply the individual total effect averaged over all individuals and clusters. This contrasts with the overall effect which is simply a comparison of the proportions who develop the outcome under allocation α_1 versus under allocation α_0 . The effect partitioning result of Hudgens and Halloran (2008) shows that the total effect is the sum of a direct effect and an indirect effect where the direct effect is defined as the effect on an individual of having the vaccine in a cluster with allocation α_1 versus not having the vaccine in a cluster with allocation α_1 and the indirect effect is defined as the effect on an individual of not having the vaccine in a cluster with allocation α_1 versus not having the vaccine in a cluster with allocation α_0 .

The effect partitioning result of Hudgens and Halloran (2008) is subject to two important criticisms. First, it is an effect partitioning result for the total effect not the overall effect and it is arguably the overall effect that is of the greatest policy interest. The overall effect for example would compare infection rates under two different allocations vaccinating say 50% versus 30% of each cluster. It is thus arguably the overall effect for which one would like a partition into direct and indirect components. Second, under the partitioning result given by Hudgens and Halloran (2008), the direct effect in this partitioning doesn't carry a clear interpretation as a "direct effect." In the definition given by Hudgens and Halloran (2008), the direct effect is the effect of having the vaccine in a cluster with allocation α_1 versus not having the vaccine in a cluster with allocation α_1 . In this comparison, of the remaining individuals in the cluster other than the one in view, one fewer of these remaining individuals receives treatment in the former scenario than in the latter (so as to maintain overall allocation α_1). The "direct effect" as defined by Hudgens and Halloran (2008) thus captures both the effect of an individual's being vaccinated but also the effect of the fact that one fewer of the other individuals in the cluster was vaccinated. It is thus not entirely clear in what sense the label "direct" is justified as the vaccine/treatment status of the other individuals in the cluster are not being held fixed. Sobel (2006) and Hudgens and Halloran (2008) also consider a setting in which one of the allocations, α_0 , leaves everyone untreated and show that the overall effect can then be partitioned as the sum of the probability of being treated under α_1 times the total effect plus the probability of being untreated under α_1 times the indirect effect. However, this is not a partitioning into a direct and indirect effect.

In this paper we give new effect partitioning results. We partition the overall effect, not the total effect, and in this partitioning we show that the overall effect is the sum of an indirect effect along with a contrast between two direct effects and that these two direct effects can in fact be appropriately interpreted as "direct" effects i.e. effects for which the treatment of all other individuals in the cluster are held constant. We show that this partitioning holds

under both randomization schemes of type A and type B, albeit under slightly different definitions of the direct and indirect effect in each of these two cases.

2. Interference and the Effect Partitioning of Hudgens and Halloran (2008)

In this section we review concepts and notation for causal inference under interference as described in Hudgens and Halloran (2008) and Tchetgen Tchetgen and VanderWeele (2011) and we review the effect partitioning result of Hudgens and Halloran (2008). Suppose there are $N > 1$ clusters in the study and for $i = 1, \dots, N$, and let n_i denote the number of individuals in cluster i . Let $\mathbf{A}_i \equiv (A_{i1}, \dots, A_{in_i})$ denote the treatments received by individuals in cluster i where A_{ij} is a dichotomous random variable. Let $\mathbf{A}_{i,-j} = (A_{i1}, \dots, A_{ij-1}, A_{ij+1}, \dots, A_{in_i})$ denote the treatments received by individuals in cluster i other than that of individual j . We refer to $\mathbf{A}_i$ as a treatment or allocation program, to distinguish it from the individual treatment A_{ij} . Let $\mathcal{A}(n)$ be the set of possible treatment allocations of length n ; thus $\mathbf{A}_i$ takes one of 2^{n_i} possible values in $\mathcal{A}(n_i)$. For each cluster i , we assume there exist counterfactual (potential outcome) data $\mathbf{Y}_i(\cdot) = \{\mathbf{Y}_i(\mathbf{a}_i) : \mathbf{a}_i \in \mathcal{A}(n_i)\}$ where $\mathbf{Y}_i(\mathbf{a}_i) = \{Y_{i1}(\mathbf{a}_i), \dots, Y_{in_i}(\mathbf{a}_i)\}$, and $Y_{ij}(\mathbf{a}_i)$ is the outcome for individual j in cluster i had treatment allocation in cluster i been $\mathbf{a}_i$. The observed outcome Y_{ij} for individual j in cluster i is equal to the potential outcome $Y_{ij}(\mathbf{A}_i)$ under the realized treatment allocation $\mathbf{A}_i$. Because $Y_{ij}(\mathbf{a}_i)$ may depend on the entire vector $\mathbf{a}_i$, we allow for the outcome of one individual to depend on the treatment received by other individuals in the same cluster i.e. we allow for interference. We assume that counterfactuals for an individual in cluster i do not depend on treatment assignments of individuals in a different cluster $i' \neq i$ (Sobel, 2006; Hudgens and Halloran, 2008). 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 $Y_{ij}(\mathbf{a}_i) = Y_{ij}(\mathbf{a}'_i)$. We suppose that $\mathbf{Y}_i(\cdot)$ is fixed for each individual and treatment allocation only determine which potential outcome is observed for each individual.

In a two-stage randomized experiment evaluating two allocation programs, each cluster is first randomized to one of the two allocation programs, say α_0 and α_1 , and then within each cluster, individuals are randomized to receive treatment with probability determined by the allocation their clustered was assigned. We let $\pi_i(\mathbf{A}_i; \alpha)$ denote that probability that treatment in cluster i takes value $\mathbf{A}_i$ under allocation program α . Following Tchetgen Tchetgen and VanderWeele (2011), we will consider two different randomization parameterizations within each cluster as given in the next definition.

Definition 1: (A) A parametrization is said to be of type A with parameter $\alpha = (n_i, K_{0,i})$ for cluster i if the treatment program $\mathbf{A}_i$ in cluster i is randomly allocated conditional on

$\mathbf{1}_n^T \mathbf{A}_i = \sum_{j=1}^{n_i} A_{ij} = K_{0,i}$ with probability mass function

$$\pi_i(\mathbf{A}_i; \alpha) = I(A_i \in A(n_i, K_{0,i})) / \binom{n_i}{K_{0,i}}$$

(B) A parametrization is said to be of type B with parameter α if treatment program $\mathbf{A}_i$ is randomly assigned to individuals in cluster i according to the known Bernoulli probability mass function

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

where $0 < \alpha < 1$.

A type A parameterization might assign half the subjects in a cluster to $A_{ij} = 1$ and half to $A_{ij} = 0$. A type B parameterization in contrast might independently assign each individual $A_{ij} = 1$ with probability one half and $A_{ij} = 0$ with probability one half so that there is a non-negative probability that treatment is given to all individuals in the cluster (or to none). Throughout we will use α to reference a specific allocation (e.g. assign 50% of individuals in the cluster the vaccine or 30% of individuals the vaccine) and we will also use α to denote the proportion vaccinated (e.g. 50% or 30%). Further below, we will also denote probability distributions arising from a specific allocation α by $\Pr_\alpha$.

Hudgens and Halloran (2008) considered comparing two treatment allocation programs of type A e.g. under allocation programs α_0 and α_1 . Building on the work of Halloran and Struchiner (1995) and Sobel (2006), Hudgens and Halloran (2008) define the direct, indirect, total and overall effects for type A parameterizations. So as to better correspond with prior causal inference literature we reverse the roles played by 0 and 1 in Hudgens and Halloran (2008). However, none of the substantive points made below are affected by doing so. They define the individual direct effect for individual j in group i when all other individuals in the group receive treatment $\mathbf{a}_{i,-j}$ by $DE_{ij}(\mathbf{a}_{i,-j}) = Y_{ij}(\mathbf{a}_{i,-j}, a_{ij} = 1) - Y_{ij}(\mathbf{a}_{i,-j}, a_{ij} = 0)$; the individual indirect effect comparing treatment allocation $\mathbf{a}_i$ and

$\mathbf{a}'_i$ by $IE_{ij}(\mathbf{a}_{i,-j}, \mathbf{a}'_{i,-j}) = Y_{ij}(\mathbf{a}_{i,-j}, a_{ij} = 0) - Y_{ij}(\mathbf{a}'_{i,-j}, a_{ij} = 0)$; the individual total effect by

$TE_{ij}(\mathbf{a}_{i,-j}, \mathbf{a}'_{i,-j}) = Y_{ij}(\mathbf{a}_{i,-j}, a_{ij} = 1) - Y_{ij}(\mathbf{a}'_{i,-j}, a_{ij} = 0)$; and the individual overall effect by

$OE_{ij}(\mathbf{a}_i, \mathbf{a}'_i) = Y_{ij}(\mathbf{a}_i) - Y_{ij}(\mathbf{a}'_i)$. Sobel (2006) refers to the indirect effect as a "spillover effect."

Note that under the assumption of no interference between individuals in a group, the individual indirect effect is equal to zero. Hudgens and Halloran (2008) also give corresponding definitions for the individual average direct, total and overall effects as stated in Definition 2 below. The superscript H in the notation given will denote definitions provided by Hudgens and Halloran so that we can distinguish this notation from similar definitions given in the next section.

Definition 2 (*Individual Average Direct, Indirect, Total and Overall Effects of Hudgens and Halloran (2008)*): Let

$$\begin{aligned} \bar{Y}_{ij}^H(a; \alpha) &= \sum_{s \in \mathcal{S}(n-1)} Y_{ij}(\mathbf{a}_{i,-j} = \mathbf{s}, a_{ij} = a) \Pr_\alpha(\mathbf{A}_{i,-j} = \mathbf{s} | A_{ij} = a) \\ \text{where } \Pr_\alpha(\mathbf{A}_{i,-j} = \mathbf{s} | A_{ij} = a) &= \frac{\pi_i(\mathbf{A}_{i,-j} = \mathbf{s}, A_{ij} = a; \alpha)}{\sum_{s' \in \mathcal{S}(n-1)} \pi_i(\mathbf{A}_{i,-j} = \mathbf{s}', A_{ij} = a; \alpha)} \end{aligned}$$

and define

$$\begin{aligned} \overline{DE}_{ij}^H(\alpha) &= \bar{Y}_{ij}^H(1; \alpha) - \bar{Y}_{ij}^H(0; \alpha) \\ \overline{IE}_{ij}^H(\alpha_1, \alpha_0) &= \bar{Y}_{ij}^H(0; \alpha_1) - \bar{Y}_{ij}^H(0; \alpha_0) \\ \overline{TE}_{ij}^H(\alpha_1, \alpha_0) &= \bar{Y}_{ij}^H(1; \alpha_1) - \bar{Y}_{ij}^H(0; \alpha_0). \end{aligned}$$

Finally let

$$\bar{Y}_{ij}^H(\alpha) = \sum_{s \in \mathcal{S}(n)} Y_{ij}(\mathbf{a}_i = \mathbf{s}) \pi_i(\mathbf{A}_i = \mathbf{s}; \alpha)$$

and define $\overline{OE}_{ij}^H(\alpha_1, \alpha_0) = \bar{Y}_{ij}^H(\alpha_1) - \bar{Y}_{ij}^H(\alpha_0)$.

Note that individual *average* direct, indirect, total and overall effects are defined with respect to comparing two allocations, α_1 and α_0 and are averaged over all possible treatment assignments under these allocations. Hudgens and Halloran (2008) go on to define the group

average direct effect by $\overline{DE}_i^H(\alpha) = \sum_{j=1}^{n_i} \overline{DE}_{ij}^H(\alpha) / n_i$ and the population average direct effect

by $\overline{DE}^H(\alpha) = \sum_{i=1}^N \overline{DE}_i^H(\alpha) / N$; the group average indirect effect as

$\overline{IE}_i^H(\alpha_1, \alpha_0) = \sum_{j=1}^{n_i} \overline{IE}_{ij}^H(\alpha_1, \alpha_0) / n_i$ and the population average indirect effect as

$\overline{IE}^H(\alpha_1, \alpha_0) = \sum_{i=1}^N \overline{IE}_i^H(\alpha_1, \alpha_0) / N$; the group average total effect by

$\overline{TE}_i^H(\alpha_1, \alpha_0) = \sum_{j=1}^{n_i} \overline{TE}_{ij}^H(\alpha_1, \alpha_0) / n_i$ and the population average total effect by

$\overline{TE}^H(\alpha_1, \alpha_0) = \sum_{i=1}^N \overline{TE}_i^H(\alpha_1, \alpha_0) / N$; and finally, the group average overall effect by

$\overline{OE}_i^H(\alpha_1, \alpha_0) = \sum_{j=1}^{n_i} \overline{OE}_{ij}^H(\alpha_1, \alpha_0) / n_i$ and the population average overall effect by

$\overline{OE}^H(\alpha_1, \alpha_0) = \sum_{i=1}^N \overline{OE}_i^H(\alpha_1, \alpha_0) / N$.

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, $\overline{TE}_{ij}^H(\alpha_1, \alpha_0) = \overline{DE}_{ij}^H(\alpha_1) + \overline{IE}_{ij}^H(\alpha_1, \alpha_0)$ (Hudgens and Halloran, 2008). Note that if interference is absent then the indirect effect, $\overline{IE}_{ij}^H(\alpha_1, \alpha_0)$, is zero, and the total and direct effects coincide and are equal to a comparison of vaccinating everyone versus vaccinating no one.

Let $S \equiv (S_1, \dots, S_N)$ denote the first stage of randomization group assignments with $S_i = 1$ if cluster i is assigned to α_1 and 0 if cluster i is assigned to α_0 . Let η denote the parametrization for the distribution of S and suppose that in a two-stage randomized experiment $\{\eta, \alpha_0, \alpha_1\}$ are Type A parametrizations.

$$\begin{aligned}
\widehat{Y}_i(a; \alpha) &= \sum_{j=1}^{n_i} I(A_{ij}) \\
&= a) Y_{ij}(\mathbf{A}_i) \\
&/ \sum_{j=1}^{n_i} I(A_{ij}) \\
&= a), \widehat{Y}(a; \alpha) \\
&= \sum_{i=1}^N \widehat{Y}_i(a; \alpha) I(S_i) \\
&= 1) / \sum_{i=1}^N I(S_i) \\
&= 1), \widehat{Y}_i(\alpha) \\
&= \sum_{j=1}^{n_i} Y_{ij}(\mathbf{A}_i) \\
&/ n_i, \text{ and } \widehat{Y}(\alpha) \\
&= \sum_{i=1}^{n_i} \widehat{Y}_i(\alpha) I(S_i) \\
&= 1) / \sum_{i=1}^N I(S_i = 1).
\end{aligned}$$

Suppose $S_i = 1$ corresponds to allocation program α , then let Hudgens and Halloran (2008) proposed the following estimators:

$$\begin{aligned}
\widehat{DE}(\alpha_0) &= \widehat{Y}(1; \alpha_0) - \widehat{Y}(0; \alpha_0), \\
\widehat{IE}(\alpha_1, \alpha_0) &= \widehat{Y}(0; \alpha_1) - \widehat{Y}(0; \alpha_0), \\
\widehat{TE}(\alpha_1, \alpha_0) &= \widehat{Y}(1; \alpha_1) - \widehat{Y}(0; \alpha_0), \\
\widehat{OE}(\alpha_1, \alpha_0) &= \widehat{Y}(\alpha_1) - \widehat{Y}(\alpha_0),
\end{aligned}$$

which they showed to be unbiased i.e.

$E\{\widehat{DE}(\alpha_0)\} = \overline{DE}^H(\alpha_0)$, $E\{\widehat{IE}(\alpha_1, \alpha_0)\} = \overline{IE}^H(\alpha_1, \alpha_0)$, $E\{\widehat{TE}(\alpha_1, \alpha_0)\} = \overline{TE}^H(\alpha_1, \alpha_0)$, $E\{\widehat{OE}(\alpha_1, \alpha_0)\} = \overline{OE}^H(\alpha_1, \alpha_0)$, where the expectation is taken with respect to the joint density of $(\mathbf{S}, \mathbf{A}_1, \dots, \mathbf{A}_N)$. Tchetgen Tchetgen and VanderWeele (2011) provided finite sample confidence intervals for these estimators.

As noted above, it follows from the definitions of Hudgens and Halloran (2008), that total effects at the individual, group or population levels can be partitioned into the sum of direct and indirect causal effects at the corresponding level. That is, for example,

$\overline{TE}_{ij}^H(\alpha_1, \alpha_0) = \overline{DE}_{ij}^H(\alpha_1) + \overline{IE}_{ij}^H(\alpha_1, \alpha_0)$. This effect partitioning is subject to two potential criticisms. First, the partitioning is for the total effect rather than for the overall effect. The overall effect is the comparison of the expected proportion with the outcome under allocation α_1 versus the expected proportion with the outcome under allocation α_0 . It is the arguably the most important policy effect of interest. It is the effect of interest that compares the infection rates that would arise under two different proportions allocated (e.g. 50% versus 30%). The group average total effect in contrast compares a scenario in which an individual is treated in a cluster with allocation α_1 to a scenario in which the individual is not treated in a cluster with allocation α_0 and then averages this contrast over all individuals in the cluster. This total effect contrast is of less interest insofar as if α_1 and α_0 are allocations in which some individuals do and some do not receive treatment, then it is impossible that each individual receives the treatment under α_1 as considered in first scenario concerning the total effect. The total effect thus, while of some interest from a hypothetical standpoint, is of much less interest from a policy perspective.

The second potential criticism of the effect partitioning of Hudgens and Halloran (2008) is that it is not entirely clear that what they defined as a direct effect,

$\overline{DE}_{ij}^H(\alpha) = \overline{Y}_{ij}^H(1; \alpha) - \overline{Y}_{ij}^H(0; \alpha)$, actually merits the label "direct effect." The direct effect of Hudgens and Halloran (2008) compares a scenario in which an individual is treated in a cluster with allocation α to a scenario in which the individual is not treated in a cluster with allocation α . In this second scenario, however, because the individual is not treated and the cluster has allocation α , of the remaining individuals in the cluster, one more will be treated in the second scenario than in the first scenario because in the first scenario the individual himself receives treatment. The treatment of the other individuals in the cluster are not be held fixed, even in proportion. To take an extreme scenario suppose there were just two individuals in a cluster and the allocation program α consisted of treating just one of the two

then we would have that $\overline{DE}_{i1}^H(\alpha) = Y_{i1}((1, 0)) - Y_{i1}((0, 1))$ i.e. a comparison for individual 1 of individual 1's receiving treatment (but not individual 2) versus individual 2's receiving treatment (but not individual 1). This effect may be of interest but it is not clear it merits the label "direct effect" because the treatments of both individuals are being changed in the comparison being made. What is effectively being held "fixed" in the direct effect of Hudgens and Halloran (2008) is effectively the allocation program, not the treatments received by other individuals.

In the next section we provide new effect partitioning results to remedy both of these potential criticisms. We give an effect partitioning result for the overall effect (rather than the total effect) and moreover the direct effect within the partitioning truly does have an interpretation of a "direct effect."

3. New Effect Partitioning Results for the Overall Effect

We begin with an effect partitioning result for two-stage randomized experiments under Type B parameterizations wherein each individual has a fixed probability of receiving treatment independent of the treatments received by other individuals in the cluster. We then show a similar result in fact holds also for two-stage randomized experiments under Type A parameterizations. We first define the effect of interest under type B parameterizations.

Definition 3 (*Individual Average Direct, Indirect, Total and Overall Effects Under Type B Parameterizations*): Let

$$\begin{aligned}\overline{Y}_{ij}^B(a; \alpha) &= \sum_{\mathbf{s} \in \mathcal{S}'(n-1)} Y_{ij}(\mathbf{a}_{i,-j}=\mathbf{s}, a_{ij}=a) \Pr_{\alpha}(\mathbf{A}_{i,-j}=\mathbf{s}) \\ \overline{Y}_{ij}^B(\alpha) &= \sum_{\mathbf{s} \in \mathcal{S}'(n)} Y_{ij}(\mathbf{a}_i=\mathbf{s}) \Pr_{\alpha}(\mathbf{A}_i=\mathbf{s})\end{aligned}$$

$$\text{where } \Pr_{\alpha}(\mathbf{A}_i=\mathbf{s}) = \prod_{j=1}^{n_i} \alpha^{A_{ij}} (1-\alpha)^{1-A_{ij}} \text{ and define}$$

$$\begin{aligned}\overline{DE}_{ij}^B(\alpha) &= \overline{Y}_{ij}^B(1; \alpha) - \overline{Y}_{ij}^B(0; \alpha) \\ \overline{IE}_{ij}^B(\alpha_1, \alpha_0) &= \overline{Y}_{ij}^B(0; \alpha_1) - \overline{Y}_{ij}^B(0; \alpha_0) \\ \overline{OE}_{ij}^B(\alpha_1, \alpha_0) &= \overline{Y}_{ij}^B(\alpha_1) - \overline{Y}_{ij}^B(\alpha_0).\end{aligned}$$

Note that for the effects in Definition 3, in contrast with those in Definition 2, the potential outcomes are averaged over the unconditional distribution of $\mathbf{A}_i$ whereas in Definition 2 this averaging takes place over the conditional distribution of $\mathbf{A}_{i,-j}$ given $A_{ij} = a$.

Group average ($\overline{DE}_i^B(\alpha)$, $\overline{IE}_i^B(\alpha_1, \alpha_0)$, $\overline{OE}_i^B(\alpha_1, \alpha_0)$) and population average ($\overline{DE}^B(\alpha)$, $\overline{IE}^B(\alpha_1, \alpha_0)$, $\overline{OE}^B(\alpha_1, \alpha_0)$) direct, indirect and overall effects could be defined by averaging over individuals and then over clusters respectively. Suppose $S_i = 1$ corresponds

$$\begin{aligned}\widehat{Y}_i(a; \alpha) &= \sum_{j=1}^{n_i} I(A_{ij}) \\ &= a)Y_{ij}(\mathbf{A}_i) \\ &/ \sum_{j=1}^{n_i} I(A_{ij}) \\ &= a), \widehat{Y}_i(a; \alpha) \\ &= \sum_{i=1}^N \widehat{Y}_i(a; \alpha) I(S_i) \\ &= 1) / \sum_{i=1}^N I(S_i) \\ &= 1), \widehat{Y}_i(\alpha) \\ &= \sum_{j=1}^{n_i} Y_{ij}(\mathbf{A}_i) \\ &/ n_i, \text{ and } \widehat{Y}(\alpha) \\ &= \sum_{i=1}^N \widehat{Y}_i(\alpha) I(S_i) \\ &= 1) / \sum_{i=1}^N I(S_i = 1). \text{ It is then straightforward to show}\end{aligned}$$

to allocation program α , and let $\widehat{Y}(1; \alpha) - \widehat{Y}(0; \alpha)$, $\widehat{Y}(0; \alpha_1) - \widehat{Y}(0; \alpha_0)$ and $\widehat{Y}(\alpha_1) - \widehat{Y}(\alpha_0)$ are unbiased for the population average direct, indirect and overall effects respectively.

Note that for two-stage randomized experiments under Type B parameterizations, the direct effect in Definition 3 defined by $\overline{DE}_{ij}^B(\alpha) = \overline{Y}_{ij}^B(1; \alpha) - \overline{Y}_{ij}^B(0; \alpha)$, now merits the label "direct effect." This direct effect, $\overline{DE}_{ij}^B(\alpha)$, compares what would happen to an individual with versus without treatment where in both scenarios every other individual received treatment with probability α . Our definition of the direct effect under parameterization B thus circumvents one of the two criticisms of the effect partitioning result of Hudgens and Halloran (2008). Our next result shows that the other criticism is circumvented as well.

Theorem 1 (Overall Effect Partitioning Under Type B Parameterization). Suppose $\pi_i(\mathbf{A}_i; \alpha_0)$ and $\pi_i(\mathbf{A}_i; \alpha_1)$ are type B parameterizations then

$$\overline{OE}_{ij}^B(\alpha_1, \alpha_0) = \alpha_1 \overline{DE}_{ij}^B(\alpha_1) - \alpha_0 \overline{DE}_{ij}^B(\alpha_0) + \overline{IE}_{ij}^B(\alpha_1, \alpha_0).$$

Theorem 1 partitions the overall effect, $\overline{OE}_{ij}^B(\alpha_1, \alpha_0)$, into two components: (i) the indirect effect, $\overline{IE}_{ij}^B(\alpha_1, \alpha_0)$, which carries with it the interpretation of comparing the outcome of an individual without treatment where every other other individual received treatment with probability α_1 versus with probability α_0 and (ii) a contrast of two direct effects, under allocation α_1 and under allocation α_0 each multiplied by the probability of an individual receiving treatment under that allocation i.e. $\alpha_1 \overline{DE}_{ij}^B(\alpha_1) - \alpha_0 \overline{DE}_{ij}^B(\alpha_0)$. The contrast,

$\alpha_1 \overline{DE}_{ij}^B(\alpha_1) - \alpha_0 \overline{DE}_{ij}^B(\alpha_0)$, is in many ways intuitive. Under allocation α_0 , each individual will have treatment with probability α_0 and thus the direct effect, $\overline{DE}_{ij}^B(\alpha_0)$, will be in operation with probability α_0 . Likewise, under allocation α_1 , each individual will have treatment with probability α_1 and thus the direct effect, $\overline{DE}_{ij}^B(\alpha_1)$, will be in operation with probability α_1 . For an overall change in allocation from α_0 to a more extensive allocation α_1 , everyone will benefit from the indirect effect, $\overline{IE}_{ij}^B(\alpha_1, \alpha_0)$, of there being a larger proportion of individuals treated and whereas a proportion α_0 benefited from the direct effect $\overline{DE}_{ij}^B(\alpha_0)$ under allocation α_0 , but now a proportion α_1 will benefit from the direct effect of treatment, $\overline{DE}_{ij}^B(\alpha_1)$, under allocation α_1 .

In fact, a similar decomposition holds for two-stage randomized experiments under parameterizations of type A. We first introduce alternative notions of direct and indirect effects under type A parameterization to those given by Hudgens and Halloran (2008).

Definition 4 (*Alternative Individual Average Direct and Indirect Effects Under Type A Parameterizations*): Let

$$\begin{aligned}\bar{Y}_{ij}^A(a; \alpha, a^*) &= \sum_{s \in \mathcal{S}(n-1)} Y_{ij}(\mathbf{a}_{i,-j} = \mathbf{s}, a_{ij} = a) \Pr_{\alpha}(\mathbf{A}_{i,-j} = \mathbf{s} | a_{ij} = a^*) \\ \bar{Y}_{ij}^A(a; \alpha) &= \sum_{s \in \mathcal{S}(n)} Y_{ij}(\mathbf{a}_{i,-j} = \mathbf{s}, a_{ij} = a) \Pr_{\alpha}(\mathbf{A}_i = \mathbf{s}) \\ \bar{Y}_{ij}^A(\alpha) &= \sum_{s \in \mathcal{S}(n)} Y_{ij}(\mathbf{a}_i = \mathbf{s}) \Pr_{\alpha}(\mathbf{A}_i = \mathbf{s})\end{aligned}$$

and define

$$\begin{aligned}\overline{DE}_{ij}^A(\alpha, a) &= \bar{Y}_{ij}^A(1; \alpha, a) - \bar{Y}_{ij}^A(0; \alpha, a) \\ \overline{IE}_{ij}^A(\alpha_1, \alpha_0) &= \bar{Y}_{ij}^A(0; \alpha_1) - \bar{Y}_{ij}^A(0; \alpha_0) \\ \overline{OE}_{ij}^A(\alpha_1, \alpha_0) &= \bar{Y}_{ij}^A(\alpha_1) - \bar{Y}_{ij}^A(\alpha_0).\end{aligned}$$

Once again, group average ($\overline{DE}_i^A(\alpha, a)$, $\overline{IE}_i^A(\alpha_1, \alpha_0)$, $\overline{OE}_i^A(\alpha_1, \alpha_0)$) and population average ($\overline{DE}^A(\alpha, a)$, $\overline{IE}^A(\alpha_1, \alpha_0)$, $\overline{OE}^A(\alpha_1, \alpha_0)$) direct, indirect and overall effects could be defined by averaging over individuals and then clusters respectively. As with Definition 3, the definition of a direct effect given in Definition 4, again merits the label "direct effect." The

direct effect in Definition 4, $\overline{DE}_{ij}^A(\alpha, a) = \bar{Y}_{ij}^A(1; \alpha, a) - \bar{Y}_{ij}^A(0; \alpha, a)$, compares the outcome for an individual with versus without treatment where in *both* scenarios all other individuals are assigned treatment under allocation α conditional on the individual in question having been given treatment a . This contrasts with the "direct effect" of Hudgens and Halloran in which the allocation rule for the other individuals in the cluster varied according to whether or not the individual in question received treatment. The indirect effect in Definition 4 is also subtly different than that of Hudgens and Halloran (2008). In the indirect effect in Definition 4, the individual is untreated in both scenarios i.e. under allocation program α_1 and α_0 but other individuals in the cluster are assigned treatment under allocations α_1 or α_0 without conditioning on the fact that the individual in question was untreated. As shown in the following theorem, the same effect partitioning effectively holds for the effects given in

Definition 4 under type A parameterization as for those given in Definition 3 under type B parameterizations.

Theorem 2. (*Overall Effect Partitioning Under Type A Parameterization*). Suppose $\pi_i(\mathbf{A}_i; \alpha_0)$ and $\pi_i(\mathbf{A}_i; \alpha_1)$ are type A parameterizations with $\alpha_1 = (n_i, K_{1,i})$ and $\alpha_0 = (n_i, K_{0,i})$ then

$$\overline{OE}_{ij}^A(\alpha_1, \alpha_0) = \frac{K_{1,i}}{n_i} \overline{DE}_{ij}^A(\alpha_1, 1) - \frac{K_{0,i}}{n_i} \overline{DE}_{ij}^A(\alpha_0, 1) + \overline{IE}_{ij}^A(\alpha_1, \alpha_0).$$

Note that under allocation α_1 , the proportion receiving treatment is simply $\frac{K_{1,i}}{n_i}$ and under

allocation α_0 , the proportion receiving treatment is simply $\frac{K_{0,i}}{n_i}$. The effect partitioning result in Theorem 2 for type A parameterization is thus entirely analogous to that in Theorem 1 for type B parameterization. Under type B parameterization however we were able to identify the population average direct, indirect and overall effects defined in Definition 3 and partitioned in Theorem 1. For type A parameterizations, the population average direct and indirect effects given in Definition 4 and partitioned in Theorem 2 are not identified. This is

because, for example, population averages of $\overline{Y}_{ij}^A(1; \alpha, 0)$ are not identified because for no cluster is there an individual j with treatment 1 in a cluster wherein all other individuals are assigned treatment under an allocation rule α conditional on individual j having been untreated; there are no clusters with this treatment arrangement because such a cluster would not in fact qualify as having been assigned allocation α . It may, however, be possible to identify the direct and indirect effects in Definition 4 under a modified type A two-stage randomized trial wherein randomization was extended to three stages so that those assigned to allocation α were further randomized either to actual allocation α or to modified allocation α with an individual randomly selected to receive treatment and all others allocation assigned treatment under an allocation rule α conditional on individual j having been untreated. Such developments and considerations of alternative study designs will be left for future work.

Note that in a limit scenario in which each cluster effectively had infinite sample size, the direct and indirect effects in Definition 4 would essentially reduce to those of Hudgens and Halloran (2008) in Definition 2 because $\Pr_\alpha(\mathbf{A}_{i,-j}|A_{ij}) \approx \Pr_\alpha(\mathbf{A}_{i,-j})$ and these would then be identified by the estimators given in the previous section. A rigorous formalization of this in terms of limit arguments lies beyond the scope of this paper and will be left to future work. It is, however, the case that even if one retains the definitions of Hudgens and Halloran (2008) in Definition 2, essentially the same effect partitioning holds, as is stated in our final theorem.

Theorem 3. (*Overall Effect Partitioning Under Parameterization A for the Direct and Indirect Effects of Hudgens and Halloran (2008)*). Suppose $\pi_i(\mathbf{A}_i; \alpha_0)$ and $\pi_i(\mathbf{A}_i; \alpha_1)$ are type A parameterizations with $\alpha_1 = (n_i, K_{1,i})$ and $\alpha_0 = (n_i, K_{0,i})$ then

$$\overline{OE}_{ij}^H(\alpha_1, \alpha_0) = \frac{K_{1,i}}{n_i} \overline{DE}_{ij}^H(\alpha_1) - \frac{K_{0,i}}{n_i} \overline{DE}_{ij}^H(\alpha_0) + \overline{IE}_{ij}^H(\alpha_1, \alpha_0).$$

Here the population average direct and indirect effects, $\overline{DE}^H(\alpha_1)$ and $\overline{IE}^H(\alpha_1, \alpha_0)$, are identified under simple type A two-stage randomized experiments and the effect partitioning holds but the direct effect does not carry a clear interpretation as a "direct" effect.

Theorems 1, 2 and 3 taken together show that the result partitioning the overall effect into the sum of a indirect effect and a contrast between two direct effects thus applies to the effects defined in Hudgens and Halloran (under type A allocations) and also to those given in this paper in Definition 4 (for type A allocations) and Definition 3 (for type B allocations).

4. Illustration

Consider the hypothetical data in Table 1 presented by Hudgens and Halloran (2008), based on a hypothetical trial similar to that of Ali et al. (2005) of a cholera vaccine, where 30% are allocated treatment under α_0 ($S_i = 0$) and 50% are allocated treatment under α_1 ($S_i = 1$).

Suppose this had in fact come from a two-stage randomized experiment under type B parameterization. Our estimates of the population average direct, indirect and overall effects in Definition 3 per 1,000 individuals would then be

$\overline{DE}^B(\alpha_0) = -3.64$, $\overline{DE}^B(\alpha_1) = -1.3$, $\overline{IE}^B(\alpha_1, \alpha_0) = -2.81$, $\overline{OE}^B(\alpha_1, \alpha_0) = -2.37$. We would then have the decomposition of the overall effect as

$$\begin{aligned}\overline{OE}^B(\alpha_1, \alpha_0) &= \alpha_1 \overline{DE}^B(\alpha_1) - \alpha_0 \overline{DE}^B(\alpha_0) + \overline{IE}^B(\alpha_1, \alpha_0) \\ \text{i.e. } -2.37 &= (0.5)(-1.3) - (0.3)(-3.64) + (-2.81).\end{aligned}$$

If the data had come from a two-stage randomized experiment under type A parameterization then, as in Hudgens and Halloran, the population average direct, indirect, total and overall effects in Definition 2 per 1,000 individuals would then be

$\overline{DE}^H(\alpha_0) = -3.64$, $\overline{DE}^H(\alpha_1) = -1.3$, $\overline{IE}^H(\alpha_1, \alpha_0) = -2.81$, $\overline{TE}^H(\alpha_1, \alpha_0) = -4.11$, $\overline{OE}^H(\alpha_1, \alpha_0) = -2.37$. Hudgens and Halloran (2008) note that

$$\begin{aligned}\overline{TE}^H(\alpha_1, \alpha_0) &= \overline{DE}^H(\alpha_1) + \overline{IE}^H(\alpha_1, \alpha_0) \\ \text{i.e. } -4.11 &= (-1.3) + (-2.81).\end{aligned}$$

Hudgens and Halloran comment that an estimate of direct effect of the vaccine of -1.3 if all clusters had been given a 50% allocation would underestimate the total effect of -4.11 considerably.

We note further here, illustrating Theorem 3, that

$$\begin{aligned}\overline{OE}^H(\alpha_1, \alpha_0) &= \alpha_1 \overline{DE}^H(\alpha_1) - \alpha_0 \overline{DE}^H(\alpha_0) + \overline{IE}^H(\alpha_1, \alpha_0) \\ \text{i.e. } -2.37 &= (0.5)(-1.3) - (0.3)(-3.64) + (-2.81).\end{aligned}$$

Here we see that the overall effect in fact is smaller in magnitude than the indirect effect. This is because the direct effect is reduced from -3.64 when only 30% are given the vaccine to -1.3 when 50% are given the vaccine. Thus under the allocation of 50%, although a greater proportion are vaccinated, this is offset by a reduced direct effect for the vaccine resulting in a net causative contribution for the direct effect contrast and consequently we see that the overall effect is smaller than the indirect effect.

5. Discussion

In the previous example, the overall effect was smaller than the indirect effect; this was because the net contribution of the direct effect contrast was causative; by increasing the numbers vaccinated herd immunity was effectively granted and the direct effect was rendered smaller in magnitude. More generally we might refer to contrasts of the form $\alpha_1 \overline{DE}^A(\alpha_1) - \alpha_0 \overline{DE}^A(\alpha_0)$ or $\alpha_1 \overline{DE}^B(\alpha_1) - \alpha_0 \overline{DE}^B(\alpha_0)$ or $\alpha_1 \overline{DE}^H(\alpha_1) - \alpha_0 \overline{DE}^H(\alpha_0)$, as the "net direct effects contribution"; this may be in either direction. For allocation proportions α_1 that are greater than allocation proportions under α_0 , we might consider three scenarios: one in which $\overline{DE}^A(\alpha_1) < \overline{DE}^A(\alpha_0)$ and consequently the net contribution of the direct effects is protective; one in which $\overline{DE}^A(\alpha_1) > \overline{DE}^A(\alpha_0)$ but the increase in proportion treatment from α_0 to α_1 is sufficient to offset the diminished direct effect so that the net contribution of the direct effects is again protective; and finally, one in which $\overline{DE}^A(\alpha_1) > \overline{DE}^A(\alpha_0)$ but the increase in proportion treatment from α_0 to α_1 is not sufficient to offset the diminished direct effect so that the net contribution of the direct effects is causative. It was this third scenario that was illustrated in the previous section. As was also seen in the illustration, however, even if the net contribution of the direct effect is causative, the overall effect of treatment may still be highly protective.

The results given in this paper have implications for study design. We have seen that if one is interested in an effect partitioning that is (i) for the overall effect, (ii) involves direct effects that are actually interpretable as direct effects and (iii) are such that all components are identifiable then it will be important to use either a two-stage randomized experiment with type B parameterizations or a modification of the type A parameterization which extends the randomization to a third stage. We have seen also that the quantity which we have referred to as the "net direct effects contribution" is an important one for understanding the overall effect comparing different allocations and for understanding whether the overall effect will be greater than or less than the indirect effect.

We note that the effect decomposition results given in this paper hold also for outcomes that are not binary. Tchetgen Tchetgen and VanderWeele (2011) gave finite sample confidence intervals for the direct, indirect, total and overall effects of Hudgens and Halloran (2008), as given in Definition 2 above, when the outcome was binary. Future work could consider such finite sample confidence intervals for the overall, direct and indirect effects given in Definitions 2, 3 and 4 above, for both binary and continuous outcomes. Future work could also formally consider how the direct and indirect effects in Definition 2 might approximate those in Definition 4 when the size of each cluster is large.

References

- Ali M, Emch M, von Seidlein L, Yunus M, Sack DA, Rao M, Holmgren J, Clemens JD. Herd immunity conferred by killed oral cholera vaccine in Bangladesh: a reanalysis. *Lancet*. 2005; 366:44–49. [PubMed: 15993232]
- Halloran ME, Struchiner CJ. Causal inference for infectious diseases. *Epidemiology*. 1995; 6:142–151. [PubMed: 7742400]
- 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.
- Hudgens MG, Halloran ME. Towards causal inference with interference. *Journal of the American Statistical Association*. 2008; 103:832–842. [PubMed: 19081744]
- Rosenbaum PR. Interference between units in randomized experiments. *Journal of the American Statistical Association*. 2007; 102:191–200.

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.

Tchetgen Tchetgen EJ, VanderWeele TJ. Estimation of causal effects in the presence of interference. Statistical Methods in Medical Research - Special Issue on Causal Inference. 2011 in press.

VanderWeele TJ. Direct and indirect effects for neighborhood-based clustered and longitudinal data. Sociological Methods and Research. 2010; 38:515–544.

Appendix: Proofs

Proof of Theorem 1. We have that $\bar{Y}_{ij}^B(\alpha)$

$$\begin{aligned} &= \alpha \sum_{s \in \mathcal{S}(n-1)} Y_{ij}(\mathbf{a}_{i,-j}=\mathbf{s}, a_{ij}=1) \Pr_{\alpha}(\mathbf{A}_{i,-j}=\mathbf{s}) + (1-\alpha) \sum_{s \in \mathcal{S}(n-1)} Y_{ij}(\mathbf{a}_{i,-j}=\mathbf{s}, a_{ij}=0) \Pr_{\alpha}(\mathbf{A}_{i,-j}=\mathbf{s}) \\ &= \alpha[\bar{Y}_{ij}^B(1;\alpha) - \bar{Y}_{ij}^B(0;\alpha)] + \bar{Y}_{ij}^B(0;\alpha). \end{aligned}$$

$$\begin{aligned} \text{Thus, } \overline{OE}_{ij}^B(\alpha_1, \alpha_0) &= \bar{Y}_{ij}^B(\alpha_1) - \bar{Y}_{ij}^B(\alpha_0) \\ &= \alpha_1[\bar{Y}_{ij}^B(1;\alpha_1) - \bar{Y}_{ij}^B(0;\alpha_1)] - \alpha_0[\bar{Y}_{ij}^B(1;\alpha_0) - \bar{Y}_{ij}^B(0;\alpha_0)] + [\bar{Y}_{ij}^B(0;\alpha_1) - \bar{Y}_{ij}^B(0;\alpha_0)] \\ &= \alpha_1 \overline{DE}_{ij}^B(\alpha_1) - \alpha_0 \overline{DE}_{ij}^B(\alpha_0) + \overline{IE}_{ij}^B(\alpha_1, \alpha_0). \end{aligned}$$

Proof of Theorem 2. We have that

$$\begin{aligned} \bar{Y}_{ij}^A(\alpha) &= \frac{K_{0,i}}{n_i} \bar{Y}_{ij}^A(1;\alpha, 1) + \frac{n_i - K_{0,i}}{n_i} \bar{Y}_{ij}^A(0;\alpha, 0) + \left\{ \frac{K_{0,i}}{n_i} \bar{Y}_{ij}^A(0;\alpha, 1) - \frac{K_{0,i}}{n_i} \bar{Y}_{ij}^A(0;\alpha, 0) \right\} \\ &= \frac{K_{0,i}}{n_i} \{\bar{Y}_{ij}^A(1;\alpha, 1) - \bar{Y}_{ij}^A(0;\alpha, 1)\} + \frac{n_i - K_{0,i}}{n_i} \bar{Y}_{ij}^A(0;\alpha, 0) + \frac{K_{0,i}}{n_i} \bar{Y}_{ij}^A(0;\alpha, 1) \\ &= \frac{K_{0,i}}{n_i} \overline{DE}_{ij}^A(\alpha, 1) + \sum_{s \in \mathcal{S}(n-1)} Y_{ij}(\mathbf{a}_{i,-j}=\mathbf{s}, a_{ij}=0) \Pr_{\alpha}(\mathbf{A}_{i,-j}=\mathbf{s} | a_{ij}=0) \left\{ \frac{n_i - K_{0,i}}{n_i} \right\} + \sum_{s \in \mathcal{S}(n-1)} Y_{ij}(\mathbf{a}_{i,-j}=\mathbf{s}, a_{ij}=0) \Pr_{\alpha}(\mathbf{A}_{i,-j}=\mathbf{s} | a_{ij}=1) \left\{ \frac{K_{0,i}}{n_i} \right\} \\ &= \frac{K_{0,i}}{n_i} \overline{DE}_{ij}^A(\alpha, 1) + \sum_{s \in \mathcal{S}(n-1)} Y_{ij}(\mathbf{a}_{i,-j}=\mathbf{s}, a_{ij}=0) \Pr_{\alpha}(\mathbf{A}_{i,-j}=\mathbf{s}, a_{ij}=0) + \sum_{s \in \mathcal{S}(n-1)} Y_{ij}(\mathbf{a}_{i,-j}=\mathbf{s}, a_{ij}=0) \Pr_{\alpha}(\mathbf{A}_{i,-j}=\mathbf{s}, a_{ij}=1) \\ &= \frac{K_{0,i}}{n_i} \overline{DE}_{ij}^A(\alpha, 1) + \sum_{s \in \mathcal{S}(n)} Y_{ij}(\mathbf{a}_{i,-j}=\mathbf{s} \setminus s_j, a_{ij}=0) \Pr_{\alpha}(\mathbf{A}_i=\mathbf{s}) \\ &= \frac{K_{0,i}}{n_i} \overline{DE}_{ij}^A(\alpha, 1) + \bar{Y}_{ij}^A(0;\alpha). \end{aligned}$$

$$\begin{aligned} \text{Thus, } \overline{OE}_{ij}^A(\alpha_1, \alpha_0) &= \bar{Y}_{ij}^A(\alpha_1) - \bar{Y}_{ij}^A(\alpha_0) \\ &= \frac{K_{1,i}}{n_i} \overline{DE}_{ij}^A(\alpha_1, 1) + \bar{Y}_{ij}^A(0;\alpha_1) - \left\{ \frac{K_{0,i}}{n_i} \overline{DE}_{ij}^A(\alpha_0, 1) + \bar{Y}_{ij}^A(0;\alpha_0) \right\} \\ &= \frac{K_{1,i}}{n_i} \overline{DE}_{ij}^A(\alpha_1, 1) - \frac{K_{0,i}}{n_i} \overline{DE}_{ij}^A(\alpha_0, 1) + \overline{IE}_{ij}^A(\alpha_1, \alpha_0). \end{aligned}$$

Proof of Theorem 3. We have that

$$\begin{aligned} \bar{Y}_{ij}^H(\alpha) &= \frac{K_{0,i}}{n_i} \bar{Y}_{ij}^H(1;\alpha) + \frac{n_i - K_{0,i}}{n_i} \bar{Y}_{ij}^H(0;\alpha) + \left\{ \frac{K_{0,i}}{n_i} \bar{Y}_{ij}^H(0;\alpha) - \frac{K_{0,i}}{n_i} \bar{Y}_{ij}^H(0;\alpha) \right\} \\ &= \frac{K_{0,i}}{n_i} \{\bar{Y}_{ij}^H(1;\alpha) - \bar{Y}_{ij}^H(0;\alpha)\} + \frac{n_i - K_{0,i}}{n_i} \bar{Y}_{ij}^H(0;\alpha) + \frac{K_{0,i}}{n_i} \bar{Y}_{ij}^H(0;\alpha) \\ &= \frac{K_{0,i}}{n_i} \overline{DE}_{ij}^H(\alpha) + \bar{Y}_{ij}^H(0;\alpha) \end{aligned}$$

$$\begin{aligned}
\text{Thus, } \overline{OE}_{ij}^H(\alpha_1, \alpha_0) &= \overline{Y}_{ij}^H(\alpha_1) - \overline{Y}_{ij}^H(\alpha_0) \\
&= \frac{K_{1,j}}{n_i} \overline{DE}_{ij}^H(\alpha_1) + \overline{Y}_{ij}^H(0; \alpha_1) - \left\{ \frac{K_{0,j}}{n_i} \overline{DE}_{ij}^H(\alpha_0) + \overline{Y}_{ij}^H(0; \alpha_0) \right\} \\
&= \frac{K_{1,j}}{n_i} \overline{DE}_{ij}^H(\alpha_1) - \frac{K_{0,j}}{n_i} \overline{DE}_{ij}^H(\alpha_0) + \overline{IE}_{ij}^H(\alpha_1, \alpha_0).
\end{aligned}$$

Table 1

Hypothetical data from Hudgens and Halloran (2008)

Group	Assignment	Vaccinated ($A_{ij} = 1$)	Cases	Placebo ($A_{ij} = 1$)	Cases
i	S_i	Total $\sum_j A_{ij}$	$\sum_j A_{ij} Y_{ij}(A_i)$	Total $\sum_j (1 - A_{ij})$	$\sum_j (1 - A_{ij}) Y_{ij}(A_i)$
1	1	12,541	16	12,541	18
2	1	11,513	26	11,513	54
3	0	10,772	17	25,134	119
4	0	8,883	22	20,727	122
5	0	5,627	15	13,130	92