Published in final edited form as:
Int J Biostat. ; 8(2): . doi:10.2202/1557-4679.1354.

Causal Inference for Vaccine Effects on Infectiousness

M. Elizabeth Halloran
Fred Hutchinson Cancer Research Center and University of Washington

Michael G. Hudgens
University of North Carolina at Chapel Hill

Abstract

If a vaccine does not protect individuals completely against infection, it could still reduce infectiousness of infected vaccinated individuals to others. Typically, vaccine efficacy for infectiousness is estimated based on contrasts between the transmission risk to susceptible individuals from infected vaccinated individuals compared with that from infected unvaccinated individuals. Such estimates are problematic, however, because they are subject to selection bias and do not have a causal interpretation. Here, we develop causal estimands for vaccine efficacy for infectiousness for four different scenarios of populations of transmission units of size two. These causal estimands incorporate both principal stratification, based on the joint potential infection outcomes under vaccine and control, and interference between individuals within transmission units. In the most general scenario, both individuals can be exposed to infection outside the transmission unit and both can be assigned either vaccine or control. The three other scenarios are special cases of the general scenario where only one individual is exposed outside the transmission unit or can be assigned vaccine. The causal estimands for vaccine efficacy for infectiousness are well defined only within certain principal strata and, in general, are identifiable only with strong unverifiable assumptions. Nonetheless, the observed data do provide some information, and we derive large sample bounds on the causal vaccine efficacy for infectiousness estimands. An example of the type of data observed in a study to estimate vaccine efficacy for infectiousness is analyzed in the causal inference framework we developed.

Keywords

causal inference; principal stratification; interference; infectious disease; vaccine

1 Introduction

1.1 Background

Evaluating the effect of vaccination on reducing infectiousness has important public health consequences. If a vaccine does not protect well against infection, it could still substantially reduce the total number of cases if transmission from infected vaccinated individuals is reduced compared to if those individuals were not vaccinated. Interest in such effects is increasing. For instance, estimates of aspects of vaccine efficacy, including the efficacy in reducing secondary transmission, are important inputs into dynamic transmission models that evaluate the effectiveness of different intervention measures. Such models are playing a

Author Notes: M. Elizabeth Halloran, Center for Quantitative Infectious Diseases, Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, and Department of Biostatistics, University of Washington. Michael G. Hudgens, Department of Biostatistics, University of North Carolina at Chapel Hill.

growing role in vaccination policy. Thus, sound methods are needed to estimate the effects of vaccines and other interventions on infectiousness.

Typically, estimates of vaccine efficacy for infectiousness are based on contrasts between the estimated transmission risk from infected vaccinated individuals compared with the estimated transmission risk from infected unvaccinated individuals to susceptible individuals exposed within small transmission units, such as households or partnerships (Halloran et al. 1997; Halloran et al. 2003). Our goal in this paper is to develop methods to estimate the causal vaccine effects on infectiousness. Unfortunately, contrasts that condition on an event, such as infection, that occurs subsequent to receipt of vaccine or control may result in selection bias. Moreover, because the set of individuals who would become infected if vaccinated is likely not identical to those who would become infected if given control, comparisons that condition on infection, e.g., the estimated transmission risk from infected vaccinated individuals compared with the transmission risk from infected unvaccinated individuals, do not necessarily have a causal interpretation.

In this paper, we adapt the approach of principal stratification for comparing treatments adjusting for posttreatment variables (Frangakis and Rubin 2002). In the setting we consider, the treatments compared are vaccine and control, and the posttreatment variable is being infected or not. We stratify on the joint potential posttreatment infection outcomes of certain individuals in the transmission unit when vaccinated and not vaccinated. Because membership in a basic principal stratum is not a effected by whether an individual is actually assigned vaccine or control, the strata can be used in the same way as pretreatment covariates. A number of papers have adapted principal stratification to assess effects of vaccination on postinfection outcomes (Gilbert et al. 2003; Shepherd et al. 2006; Hudgens and Halloran 2006; Jemiai et al. 2007).

However, to date little has been done to develop methods of causal inference for estimating vaccine efficacy for infectiousness. The situation is further complicated, because the treatment (vaccination) status of one person in a transmission unit may affect the infection outcome in other individuals in the transmission unit, so that interference (Cox 1958; Halloran and Struchiner 1991, Hudgens and Halloran 2008; Tchetgen and VanderWeele 2010) may be present in the setting of estimating vaccine efficacy for infectiousness. Thus causal methods for estimating vaccine efficacy for infectiousness combine principal stratification with elements of interference between individuals within the transmission units. Hudgens and Halloran (2006) briefly discussed the problem of developing causal methods for estimating vaccine efficacy for infectiousness. Recently, VanderWeele and Tchetgen (2011) proposed causal estimands and bounds for vaccine effects on infectiousness.

1.2 Traditional method

Here we review a traditional method for estimating the vaccine effects on infectiousness before introducing the proposed causal estimands in the following sections. The standard method for estimating vaccine efficacy for infectiousness is based on the secondary attack rate (SAR) within small transmission units (Halloran et al. 2003). The conventional approach to estimate various vaccine effects based on the secondary attack rate assumes that the primary, i.e., first case in the transmission unit is exposed from outside the transmission unit, and the others in the transmission unit are then exposed only by the primary case during the period of observation. The proportion of exposed individuals within the transmission unit who become infected is used to estimate the secondary attack rate. If another case occurs temporally too close to the primary case to have been infected by the primary case, it is considered a co-primary, and is excluded from the analysis. To determine whether secondary transmission may have occurred, assumptions are needed about the

natural history of infection, such as the latent, incubation, and infectious periods, depending on the infectious agent under study and the ascertainment method. Genetic sequencing of the infecting pathogens can also provide information on whether secondary transmission occurred.

The SAR is defined as the observed proportion of persons within the transmission unit exposed to the primary case who become infected. Define the observed SAR from primary cases with vaccine status r to susceptible individuals with vaccine status s, with $r, s \in \{0, 1\}$ where 0 denotes control and 1 denotes vaccine, as

$$\mathrm{SAR}_{rs} = \frac{\text{no. persons with status } s \text{ infected by persons with status } r}{\text{no. of persons with status } s \text{ exposed to persons with status } r}. \quad (1)$$

For example, SAR_{01} denotes the observed proportion of vaccinated persons who became infected after exposure to a primary case who received control, and SAR_{10} denotes the observed proportion of persons who received control who became infected after exposure to a vaccinated primary case.

Halloran et al. (2003) defined three estimators of the vaccine effect on infectiousness, two stratified by the vaccine status of the exposed individuals and one not stratified by the vaccine status of the exposed individuals. The unstratified estimator is not considered further here. In general, in this paper VE denotes vaccine efficacy, VE_I denotes vaccine efficacy for infectiousness, and VE_S denotes vaccine efficacy for susceptibility. The two stratified estimators of the vaccine effect on infectiousness are based on contrasts of the secondary attack rates, SAR_{rs}, $r, s \in \{0, 1\}$. Because the secondary attack rates are estimated in transmission units where an initial person becomes infected, the contrasts are analogous to the "net" effect of Frangakis and Rubin (2002) of an assignment on the observed outcome adjusting for a posttreatment variable. The net effect compares the outcomes under two treatments in individuals with a common observed value of a posttreatment variable. In our case, the net effect of an assignment (vaccine versus control) compares the observed infection outcome in individuals exposed to individuals who actually became infected under vaccine or control. In particular, the two stratified net estimators of the vaccine effect on infectiousness are

$$\widehat{\mathrm{VE}}_I^{net}(0) = 1 - \frac{\mathrm{SAR}_{10}}{\mathrm{SAR}_{00}} \quad \text{and} \quad \widehat{\mathrm{VE}}_I^{net}(1) = 1 - \frac{\mathrm{SAR}_{11}}{\mathrm{SAR}_{01}}, \quad (2)$$

where $\widehat{\mathrm{VE}}_I^{net}(z)$ denotes the estimated stratified net vaccine efficacy for infectiousness when the exposed individuals have vaccine status z. The set of individuals who would become the primary case if vaccinated is likely not identical to those who would become the primary case if given control, so $\widehat{\mathrm{VE}}_I^{net}(0)$ and $\widehat{\mathrm{VE}}_I^{net}(1)$ could be subject to selection bias and do not necessarily have a causal interpretation.

Although it is not our primary interest here, it is helpful to define the vaccine efficacy for susceptibility when exposed to a primary case, i.e., a vaccine's ability to protect a vaccinated individual from infection when exposed in the transmission unit. Let $\widehat{\mathrm{VE}}_S^{net}(z)$ be the estimated stratified net vaccine efficacy for susceptibility when exposed to a primary case with vaccine status z, where

$$\widehat{\mathrm{VE}}_S^{net}(0) = 1 - \frac{\mathrm{SAR}_{01}}{\mathrm{SAR}_{00}} \quad \text{and} \quad \widehat{\mathrm{VE}}_I^{net}(1) = 1 - \frac{\mathrm{SAR}_{11}}{\mathrm{SAR}_{10}}. \quad (3)$$

The two stratified $\widehat{\mathrm{VE}}_S^{net}(z)$ condition on the vaccine status of the primary case and are not subject to the same potential selection bias as the stratified $\widehat{\mathrm{VE}}_I^{net}(z)$ estimators.

Using these methods, Préziosi and Halloran (2003) estimated the effects of pertussis vaccination on infectiousness and susceptibility within transmission units from a study in Niakhar, Senegal. The $\widehat{\mathrm{VE}}_I^{net}(0)$ and $\widehat{\mathrm{VE}}_I^{net}(1)$ (95% CI) were 0.63 (0.25,0.85) and 0.67 (0.29,0.87). The $\widehat{\mathrm{VE}}_S^{net}(0)$ and $\widehat{\mathrm{VE}}_S^{net}(1)$ (95% CI) were 0.31 (0.07,0.52) and 0.37 (0.09,0.60). This analysis provided some of the first direct evidence that pertussis vaccination reduces infectiousness of a clinical case of pertussis, with important policy implications. In the few instances of which we know where vaccine efficacy for infectiousness was estimated for a prophylactic vaccine, such as this pertussis vaccine example, (see also Cisse et al. (1999), Millar et al. (2008)), it has usually been in the context of an observational, or partially observational study.

The focus in this paper is the context of randomized controlled vaccine trials. Usually vaccine trials are individually randomized, but group randomized trials are also possible (Moulton et al. 2001). Vaccine trials can be quite large-scale. For example a randomized trial of two cholera vaccines with one placebo arm enrolled over 89,596 individuals (Clemens et al. 1986). A more recent pneumococcal vaccine trial enrolled 37,868 children (Black et al. 2000). However, up to now, few, if any, randomized vaccine trials have been conducted with the primary goal to estimate vaccine efficacy for infectiousness. A randomized study of how antiviral therapy could limit transmission of human immunodeficiency virus type 1 (HIV-1) in serodiscordant partners was recently conducted (Cohen et al. 2011). In that trial, however, the individuals who were randomized were already infected. Datta et al. (1999) compared the efficiency of three randomization schemes for estimating the vaccine efficacy for infectiousness and the protective efficacy in households of size two: (1) randomization by individual for a mixed allocation of vaccine and control, (2) randomization by transmission unit for a concordant allocation, and (3) randomization of only one individual in each transmission unit to either vaccine of placebo. Other theoretical aspects of estimating vaccine effects on infectiousness have been explored in several papers (Koopman and Little 1995; Rida 1996; Halloran et al. 1997; Datta et al. 1998; Becker et al. 2006).

Previous investigators have noted that studies based on pairs are well-suited to estimate vaccine effects on infectiousness as well as susceptibility (PHLS Epidemiologic Research Laboratory 1982; Koopman and Little 1995; Rida 1996; Datta et al. 1998; Becker et al. 2006). To develop methods for the causal effect of vaccination on infectiousness, in this paper we begin with the case of transmission units of size two where one or both of the individuals in the transmission unit are randomized to receive vaccine or control. As a working example, consider a randomized study in 11,000 transmission units of size two (Table 1). Individuals could be randomized to either vaccine or control, or transmission units could be randomized to one of four assignment options, in which either both receive vaccine, both receive control, or one receives vaccine and the other control, which can occur two ways. Each of the four possible combinations of vaccine and control is assigned to one quarter of the transmission units. In Table 1, 8500 transmission units have a primary case, and secondary transmission occurs in 5825 of those transmission units. The distribution of the vaccine status of the primary cases, the individuals exposed to the primary cases, and the secondary cases as well as the secondary attack rates are shown in Table 1. For these data, the two stratified net estimates of vaccine effect on infectiousness are

$$\widehat{\mathrm{VE}}_I^{net}(0) = 1 - \frac{0.45}{0.90} = 5 \quad \text{and} \quad \widehat{\mathrm{VE}}_I^{net}(1) = 1 - \frac{0.35}{0.70} = 0.5. \tag{4}$$

The estimators $\widehat{\mathrm{VE}}_S^{net}(0)$ and $\widehat{\mathrm{VE}}_S^{net}(1)$ can also be computed from Table 1.

1.3 Outline

In this paper, we extend the methods of Hudgens and Halloran (2006) for causal vaccine effects on binary postinfection outcomes within the same individual to the situation where the binary postinfection outcome is transmission from the infected person to another individual within a transmission unit. We develop causal vaccine efficacy for infectiousness estimands for four different scenarios of populations of transmission units of size two. In Section 2 we develop the most general scenario in which both individuals can be exposed to infection outside the transmission unit and both can be assigned either vaccine or control. In Section 3 identifiability arguments and large-sample bounds on the casual estimands for vaccine efficacy for infectiousness are presented. In Sections 4 through 6, three special cases of the general scenario are developed. For each scenario, we consider the basic principal stratification, identifiability, and bounds on the causal vaccine efficacy for infectiousness estimands. Section 7 discusses limitations of the approach and directions for future research.

2 Methods

2.1 Notation

Suppose we have a random sample of N transmission units of size two. Let $Z_{ij} = 1$ if individual j in transmission unit i receives vaccine and 0 otherwise for $i = 1, \ldots, N$ and $j = 1, 2$. Let $\mathbf{Z}_i = (Z_{i1}, Z_{i2})$ denote the vaccination assignment vector for transmission unit i and $Z_{i(j)}$ denote the vaccination assignment for the individual other than j in transmission unit i. Let $\mathbf{z}_i$, $z_{i(j)}$, and z_{ij} denote possible values of $\mathbf{Z}_i$, $Z_{i(j)}$, and Z_{ij}. Let the potential outcome $S_{ij}(\mathbf{z}_i) = 1$ if individual j in transmission unit i would become a primary case under treatment assignment $\mathbf{z}_i$ and $S_{ij}(\mathbf{z}_i) = 0$ otherwise. If $S_{ij}(\mathbf{z}_i) = 1$, then for $j \neq k$, let the potential outcome $Y_{ik}(\mathbf{z}_i) = 1$ if, under treatment assignment $\mathbf{z}_i$, individual k would become a secondary case infected by individual j, and $Y_{ik}(\mathbf{z}_i) = 0$ otherwise. On the other hand, if potential outcome $S_{ij}(\mathbf{z}_i) = 0$, then for $k \neq j$ we say potential outcome $Y_{ik}(\mathbf{z}_i)$ is undefined and denoted by *. That is, it does not make sense to ask whether individual k becomes a secondary case infected by individual j if individual j does not become a primary case. Let $\mathbf{S}_{i1} = (S_{i1}(00), S_{i1}(01), S_{i1}(10), S_{i1}(11))$, $\mathbf{S}_{i2} = (S_{i2}(00), S_{i2}(01), S_{i2}(10), S_{i2}(11))$, and $\mathbf{S}_i = (\mathbf{S}_{i1}, \mathbf{S}_{i2})$. Similarly, let $\mathbf{Y}_{i1} = (Y_{i1}(00), Y_{i1}(01), Y_{i1}(10), Y_{i1}(11))$, $\mathbf{Y}_{i2} = (Y_{i2}(00), Y_{i2}(01), Y_{i2}(10), Y_{i2}(11))$, and $\mathbf{Y}_i = (\mathbf{Y}_{i1}, \mathbf{Y}_{i2})$. Let $S_{ij}^{obs} = S_{ij}(\mathbf{Z}_i)$ denote the observed infection outcome depending on actual vaccine assignment, and analogously $Y_{ij}^{obs} = Y_{ij}(\mathbf{Z}_i)$ for the observed secondary transmission. Using this notation, for fixed $r, s \in \{0, 1\}$, the secondary attack rate attack $\mathrm{SAR}_{rs}(1)$ can be written

$$\mathrm{SAR}_{rs} = \frac{\sum_i \sum_{j \neq k} I\left(Z_{ij} = r, Z_{ik} = s, Y_{ik}^{obs} = 1\right)}{\sum_i \sum_{j \neq k} I\left(Z_{ij} = r, Z_{ik} = s, S_{ik}^{obs} = 1\right)}$$

where, in general, Σi denotes summation over $i = 1, \ldots, N$, $\Sigma_{j \neq k}$ denotes summation over $j, k \neq \{1, 2\}$ with $j \neq k$, and $I(\cdot)$ denotes the usual indicator, function which equals 1 if $\cdot$ is true and 0 otherwise. In the sequel it will also be helpful to define the following secondary attack rate where $j, k \in \{1, 2\}$ with $j \neq k$ also fixed:

$$\mathrm{SAR}_{rs}^{jk}=\frac{\sum_i I\left(Z_{ij}=r, Z_{ik}=s, Y_{ik}^{obs}=1\right)}{\sum_i I\left(Z_{ij}=r, Z_{ik}=s, S_{ij}^{obs}=1\right)}.$$

2.2 Assumptions and principal strata

Throughout we make the following two assumptions. First, as in Hudgens and Halloran (2008), we assume no interference across transmission units, the assumption of partial interference (Sobel 2006). This assumption is typical in conventional estimation of vaccine efficacy based on the SAR (Orenstein et al. 1988). Thus, the transmission units are assumed to be embedded in a large population such that they are transmission dynamically separated either geographically, in time, and/or socially (Halloran and Struchiner 1991). Second, we assume perfect compliance, i.e., assignment to vaccine (control) is equivalent to receipt of vaccine (control). Without this assumption, the causal estimands defined here would be causal effects of *assignment* to vaccine. The ramifications of non-compliance on the assessment of causal vaccine effects will not be considered here.

We consider first the general scenario for transmission units of size two where either individual in the transmission unit can be (i) vaccinated and (ii) exposed to infection from outside the transmission unit. Condition (i) allows for both, neither, or just one of the two individuals within a transmission unit to be vaccinated. Furthermore, the positivity assumption states that each individual has a nonzero probability of being assigned vaccine or control:

$$\text{Positivity: } 0<\Pr[\mathbf{Z}_i=\mathbf{z}_i]<1 \text{ for } \mathbf{z}_i \in \{(00),(01),(10),(11)\}. \tag{5}$$

Condition (ii) allows either of the two individuals to be a primary case. To begin, we assume the individual labels 1 and 2 are randomly assigned, thus there is no information in the labels 1 and 2. Alternatively labels 1 and 2 could indicate a covariate, such as younger versus older sibling in a household, or male versus female in partnerships. The implications of this case are considered in Section 3.7.

We define a basic principal stratification of transmission units according to the joint potential infection outcomes indicating who is a primary case under the four possible allocations of vaccine and control in the transmission unit. The vector of potential outcomes $\mathbf{S}_i$ indicating who is a primary case in transmission unit i can take on $2^8 = 256$ possible values. The number of feasible values for this vector can be decreased substantially by introducing additional assumptions. In particular, the five realistic exclusion restriction assumptions described below reduce the number of feasible values to 11. First, we assume there can be no more than one primary case for any vector of vaccination assignments $\mathbf{z}_i$, such that

$$\sum_{j=1}^{2} S_{ij}(\mathbf{z}_i) \leq 1. \tag{6}$$

This assumption implies there are no co-primary cases, where a co-primary case is an infection that occurs temporally so close to the primary case that secondary transmission could not have occurred, indicating both individuals were infected from outside the transmission unit. Assumption (6) implies that if $S_{ij}(\mathbf{z}_i) = 1$, the other person $k \neq j$ cannot be the primary case, and thus $S_{ik}(\mathbf{z}_i) = 0$. Note if person k becomes infected from outside the

transmission unit after person j becomes the primary case, then $Y_{ik}(\mathbf{z}_i) = 0$ because person k did not become infected from person j and thus would not be considered a secondary case.

Second, we assume

$$\text{If } S_{ij}\left(z_{ij}=1, z_{i(j)}=z\right)=1, \text{ then } S_{ij}\left(z_{ij}=0, z_{i(j)}=z\right)=1 \text{ for } z=0, 1, \tag{7}$$

i.e., if individual j becomes the primary case when vaccinated, then s/he will become the primary case when unvaccinated all other things being equal. Similarly, we assume

$$\text{If } S_{ij}\left(z_{ij}=z, z_{i(j)}=0\right)=1, \text{ then } S_{ij}\left(z_{ij}=z, z_{i(j)}=1\right)=1 \text{ for } z=0, 1. \tag{8}$$

We also assume

$$\text{IF } S_{ij}(00)=0 \text{ for all } j, \text{ then } S_{ij}(\mathbf{z}_i)=0 \text{ for all } j, \mathbf{z}_i. \tag{9}$$

That is, if there are no infections in a transmission unit when neither individuals is vaccinated, there cannot be any infections when one or both individuals are vaccinated. Assumptions (7) and (9) can be viewed as generalizations of the usual monotonicity assumption in the vaccine setting that the protective effect of vaccination is nonnegative (Gilbert, et al. 2003). Finally, we assume

$$\text{If there exists } j \text{ such that } S_{ij}(\mathbf{z}_i)=0 \text{ for all } \mathbf{z}_i, \text{ then } S_{ik}\left(z_{ik}=z, z_{ij}=0\right)=S_{ik}\left(z_{ik}=z, z_{ij}=1\right) \text{ for } k \neq j. \tag{10}$$

That is, if individual j is never the primary case, then this individual does not interfere with others.

Determining which of the 256 possible vectors of potential outcomes $\mathbf{S}_i$ are feasible under (6) – (10) is simple by computer. The resulting feasible 11 vectors of potential outcomes $\mathbf{S}_i$ are given in Table 2. In stratum 1, there are no primary cases in the transmission units under any of the four possible vaccination assignments. The remaining ten strata have five pairs, (2,3), (4,5), (6,9), (7,8), and (10,11), that are symmetric between individual $j = 1$ and $j = 2$. The proportion of transmission units in each stratum, denoted by $a, b, \ldots, f$, are given in Table 2. Because the labels 1 and 2 are assumed to be assigned arbitrarily to individuals, the proportion of transmission units in stratum 2 equals the proportion of transmission units in stratum 3. Similarly, the proportions of units in the symmetric pairs of strata (4,5), (6,9), (7,8), and (10,11) are equal. Because the 11 principal strata in Table 2 are exhaustive and mutually exclusive, it follows $a + 2b + 2c + 2d + 2e + 2f = 1$.

2.3 Monotonicity of Y

Each of the 11 feasible principal strata in Table 2 has an associated set of $\mathbf{Y}_i$ vectors of potential outcomes. We make an exclusion restriction assumption that vaccination has a nonnegative effect in an individual exposed within the transmission unit, analogous to the usual monotonicity assumption:

$$\text{If } Y_{ij}\left(z_{ij}=1, z_{i(j)}=z\right)=1, \text{ then } Y_{ij}\left(z_{ij}=0, z_{i(j)}=z\right)=1 \text{ for } z=0, 1. \tag{11}$$

Table 2 shows the feasible $\mathbf{Y}_{i2}$ potential outcome vectors for the feasible $\mathbf{S}_{i1}$ potential outcome vectors. In stratum 11 in which $S_{i1}(\mathbf{z}_i) = 1$ for all $\mathbf{z}_i$, the monotonicity assumption (11) reduces the types of $\mathbf{Y}_{i2}$ potential outcome vectors from $2^4 = 16$ to nine. In stratum 8, (11) reduces the number of types of transmission units from $2^3 = 8$ to six. In strata 3, 5, and

6, (11) reduces the number of types of transmission units from $2^2 = 4$ to three. The corresponding feasible $\mathbf{Y}_{i1}$ for the feasible $\mathbf{S}_{i2}$ are not shown due to space considerations.

Assumption (11) does not imply the individual causal effect of vaccination on infectiousness is non-negative. Rather, we allow that vaccination could have a negative, that is harmful, causal effect on infectiousness. As an example consider stratum 11. In one type of transmission unit, no transmission takes place regardless of vaccine allocation, and in one type, transmission always takes place. In three types of transmission unit, vaccination in the primary case enhances transmission, indicated by ↑ in Table 2. In summary, the vaccine can make an individual more infectious, but not more susceptible.

2.4 VE_I causal estimands

To define VE_I causal estimands in this setting, first consider secondary transmission events from individual 1 to individual 2. In this case, we desire a contrast of potential infection outcomes in individual 2 when (i) individual 1 is the primary case if vaccinated compared with when (ii) individual 1 is the primary case if not vaccinated. For the estimand to incorporate aspects only of VE_I, the vaccination status of individual 2 should be held fixed. These considerations suggest two estimands, one where individual 2 is not vaccinated and one where individual 2 is vaccinated. From Table 2, we see that only in stratum 11 is individual 1 the primary case under both vaccine and control when individual 2 is not vaccinated. Thus, the contrast in $Y_{i2}(10)$ and $Y_{i2}(00)$ is defined only in stratum 11. If individual 2 is vaccinated, the contrast in $Y_{i2}(11)$ and $Y_{i2}(01)$ is defined in strata 8, 9, and 11.

Let $\mathrm{CVE}_I^{jk}(z_k)$ denote the causal VE_I estimand when individual j is the primary case exposing individual k with vaccine status z_k. The first causal estimand is defined as

$$\mathrm{CVE}_I^{12}(0) = 1 - \frac{E[Y_{i2}(10) | \mathbf{S}_{i1} = (1111)]}{E[Y_{i2}(00) | \mathbf{S}_{i1} = (1111)]}. \quad (12)$$

The causal estimand is defined in the principal stratum 11 in which person $j = 1$ would become the primary case whether vaccinated or unvaccinated, and person $j = 2$ is unvaccinated. The interpretation is that the causal estimand $\mathrm{CVE}_I^{12}(0)$ is the relative reduction in the probability of secondary transmission from individual $j = 1$ to individual $j = 2$ due to vaccinating only individual $j = 1$ in the principal stratum in which person $j = 1$ would become the primary case whether vaccinated or unvaccinated. Similarly, the second estimand is

$$\mathrm{CVE}_I^{12}(1) = 1 - \frac{E[Y_{i2}(11) | \mathbf{S}_{i1} \in \{(0101), (1101), (1111)\}]}{E[Y_{i2}(01) | \mathbf{S}_{i1} \in \{(0101), (1101), (1111)\}]}. \quad (13)$$

The causal estimand is defined in the three principal strata in which person $j = 1$ would become the primary case whether vaccinated or unvaccinated, and person $j = 2$ is vaccinated. The interpretation is that the causal estimand $\mathrm{CVE}_I^{12}(1)$ is the relative reduction in the probability of secondary transmission from individual $j = 1$ to individual $j = 2$ due to vaccinating individual $j = 1$ when individual $j = 2$ is vaccinated. Alternatively, we could define the causal estimands as differences.

By symmetry analogous causal estimands can be defined for transmission from individual 2 to individual 1. Analogous to estimand (12), in stratum 10 in which individual 2 becomes infected whether vaccinated or unvaccinated and individual 1 receives control, we define

$$\mathrm{CVE}_I^{21}(0) = 1 - \frac{E[Y_{i1}(01) | \mathbf{S}_{i2} = (1111)]}{E[Y_{i1}(00) | \mathbf{S}_{i2} = (1111)]}, \tag{14}$$

i.e., the relative change in the probability of secondary transmission from individual 2 to individual 1 when individual 1 receives control. Analogous to estimand (13), in the three strata in which individual 2 becomes infected whether vaccinated or not and individual 1 is vaccinated, we define

$$\mathrm{CVE}_I^{21}(1) = 1 - \frac{E[Y_{i1}(11) | \mathbf{S}_{i2} \in \{(0011), (1011), (1111)\}]}{E[Y_{i1}(10) | \mathbf{S}_{i2} \in \{(0011), (1011), (1111)\}]}. \tag{15}$$

When individuals in a transmission unit are arbitrarily labeled 1 and 2, a single composite vaccine efficacy for infectiousness estimand can be defined. First define two new variables

$$\begin{aligned} Y_i^c(10) &= \begin{cases} Y_{i1}(01) & \text{if } \mathbf{S}_{i2} = (1111) \\ Y_{i2}(10) & \text{if } \mathbf{S}_{i1} = (1111) \end{cases} \quad \text{and} \\ Y_i^c(00) &= \begin{cases} Y_{i1}(00) & \text{if } \mathbf{S}_{i2} = (1111) \\ Y_{i2}(00) & \text{if } \mathbf{S}_{i1} = (1111) \end{cases} \end{aligned} \tag{16}$$

In words, for units in strata 10 and 11, $Y_i^c(10)$ indicates whether secondary transmission occurred from a vaccinated to unvaccinated individual, whereas $Y_i^c(00)$ indicates whether secondary transmission occurred between two unvaccinated individuals. Note in the definition of $Y_i^c(z0)$ for $z \in \{0, 1\}$ given in (16) that z denotes the vaccination status of the primary case. In contrast, in the definition of $Y_{ij}(z0)$ the z denotes the vaccination status of the individual with label 1. Using these composite variables we can define the combined estimand

$$\mathrm{CVE}_I^c(0) = 1 - \frac{E\left[Y_i^c(10) | \mathrm{S}_{i1} = (1111) \text{ or } \mathrm{S}_{i2} = (1111)\right]}{E\left[Y_i^c(00) | \mathrm{S}_{i1} = (1111) \text{ or } \mathrm{S}_{i2} = (1111)\right]}, \tag{17}$$

i.e., the relative reduction in the probability of secondary transmission to an unvaccinated individual due to vaccination of the primary case in strata 10 and 11. Similar composite variables $Y_i^c(11)$ and $Y_i^c(01)$ can be defined. Then the combined estimand is

$$\mathrm{CVE}_I^c(1) = 1 - \frac{E\left[Y_i^c(11) | \mathrm{S}_{i1} \in \{(0101), (1101), (1111)\} \text{ or } \mathrm{S}_{i2} \in \{(0011), (1011), (1111)\}\right]}{E\left[Y_i^c(01) | \mathrm{S}_{i1} \in \{(0101), (1101), (1111)\} \text{ or } \mathrm{S}_{i2} \in \{(0011), (1011), (1111)\}\right]} \tag{18}$$

i.e., the relative reduction in the risk of secondary transmission to a vaccinated individual due to vaccination.

2.5 Other estimands

The traditional estimators of vaccine efficacy for infectiousness will not in general be unbiased or consistent for the causal estimands defined in the previous section. To demonstrate this, it will be helpful to define the following estimands. The estimand for the net vaccine effect on the probability of secondary transmission from individual 1 to individual 2 when individual 2 receives control is

$$\mathrm{VE}_I^{net,12}(0) = 1 - \frac{E\left[Y_{i2}^{obs} | S_{i1}^{obs} = 1, \boldsymbol{Z}_i = (10)\right]}{E\left[Y_{i2}^{obs} | S_{i1}^{obs} = 1, \boldsymbol{Z}_i = (00)\right]}. \tag{19}$$

The estimand for the net vaccine effect on the probability of secondary transmission from individual 1 to individual 2 when individual 2 is assigned vaccine is

$$\mathrm{VE}_I^{net,12}(1) = 1 - \frac{E\left[Y_{i2}^{obs} | S_{i1}^{obs} = 1, \boldsymbol{Z}_i = (11)\right]}{E\left[Y_{i2}^{obs} | S_{i1}^{obs} = 1, \boldsymbol{Z}_i = (01)\right]}. \tag{20}$$

Analogous $\mathrm{VE}_I^{net,21}(0)$ and $\mathrm{VE}_I^{net,21}(1)$ can be defined. In settings where individuals within a transmission unit are arbitrarily labeled 1 and 2, a single composite net vaccine efficacy for infectiousness can be defined. Similar to (16), first define the new variable

$$Y_i^{c,obs} = \begin{cases} Y_{i2}^{obs} & \text{if} \quad S_{i1}^{obs} = 1 \\ Y_{i1}^{obs} & \text{if} \quad S_{i2}^{obs} = 1. \end{cases} \tag{21}$$

The combined estimand for the net vaccine effect on the risk of secondary transmission if the exposed individual receives control is

$$\mathrm{VE}_I^{net,c}(0) = 1 - \frac{E\left[Y_i^{c,obs} | \left\{S_{i1}^{obs} = 1, \boldsymbol{Z}_i = (10)\right\} \text{ or } \left\{S_{i2}^{obs} = 1, \boldsymbol{Z}_i = (01)\right\}\right]}{E\left[Y_i^{c,obs} | \left\{S_{i1}^{obs} + S_{i2}^{obs} = 1, \boldsymbol{Z}_i = (00)\right\}\right]}. \tag{22}$$

The combined estimand for the net vaccine effect on the risk of secondary transmission if the exposed individual receives vaccine is

$$\mathrm{VE}_I^{net,c}(1) = 1 - \frac{E\left[Y_i^{c,obs} | \left\{S_{i1}^{obs} + S_{i2}^{obs} = 1, \boldsymbol{Z}_i = (11)\right\}\right]}{E\left[Y_i^{c,obs} | \left\{S_{i1}^{obs} = 1, \boldsymbol{Z}_i = (01)\right\} \text{ or } \left\{S_{i2}^{obs} = 1, \boldsymbol{Z}_i = (10)\right\}\right]}. \tag{23}$$

In general, the VE_I^{net} estimands do not have a causal interpretation. For example, considering (19), as can be seen from Table 2, the set of individuals with $S_{i1}(10) = 1$ is not necessarily identical to the set of individuals with $S_{i1}(00) = 1$.

Although it is not our primary interest here, for completeness we define six analogous VE_S^{net} estimands, the net vaccine effect on protecting against infection when exposed in the transmission unit. For example, the net protective effect of vaccine in individual 2 when individual 1 receives control is

$$\mathrm{VE}_S^{net,12}(0) = 1 - \frac{E\left[Y_{i2}^{obs} | S_{i1}^{obs} = 1, \boldsymbol{Z}_i = (01)\right]}{E\left[Y_{i2}^{obs} | S_{i1}^{obs} = 1, \boldsymbol{Z}_i = (00)\right]}. \tag{24}$$

The definitions of the five $\mathrm{VE}_S^{net,12}(1)$, $\mathrm{VE}_S^{net,21}(0)$, $\mathrm{VE}_S^{net,21}(1)$, $\mathrm{VE}_S^{net,c}(0)$, and $\mathrm{VE}_S^{net,c}(0)$ estimands follow by analogy to (20), (22), and (23).

Finally, we define the causal protective effect of vaccination on infection from outside the transmission unit. We define the causal vaccine effect on infection in individual 1 by

$$\mathrm{VE}_{S,1}(0) = 1 - \frac{E[S_{i1}(10)]}{E[S_{i1}(00)]} \quad \text{and} \quad \mathrm{VE}_{S,1}(1) = 1 - \frac{E[S_{i1}(11)]}{E[S_{i1}(01)]}, \tag{25}$$

the relative average causal effect of vaccination on infection in individual 1 exposed outside the transmission unit, the first when individual 2 is not vaccinated, the second, when individual 2 is vaccinated. Four further estimands, $\mathrm{VE}_{S,2}(0)$, $\mathrm{VE}_{S,2}(1)$, $\mathrm{VE}_{S,c}(0)$ and $\mathrm{VE}_{S,c}(1)$, can be defined. These estimands are analogous to the vaccine efficacy for susceptibility, VE_S, defined by Hudgens and Halloran (2006). However, now the estimands depend on the vaccination status of the other individual in the transmission unit.

3 Identifiability

Table 3 shows the relation of the observed combinations of (S_{i1}^{obs}, S_{i2}^{obs} $\mathbf{Z}_i$) and the principal strata. Most combinations of (S_1^{obs}, S_2^{obs}, $\mathbf{Z}_i$) can belong to two or more principal strata. None of the six causal estimands (12) through (18) are identifiable from the observable data without further assumptions. We now make the assumption that vaccine assignment $\mathbf{Z}_i$ is independent of the potential outcome vector ($\mathbf{S}_i$, $\mathbf{Y}_i$) for all i, thus vaccine assignment is strongly ignorable. Formally,

$$\text{Independence:} \mathbf{Z}_i \perp \{\mathbf{S}_i, \mathbf{Y}_i\} \quad \text{for all} \quad i. \tag{26}$$

Randomization is one such assignment mechanism. Assignment to vaccine and control could be randomized individually or by transmission unit. For example, if randomized individually, then the assignments $Z_{11}, \ldots, Z_{N1}, Z_{12}, \ldots, Z_{N2}$ could be made independently with $\Pr[Z_{ij} = 11] = 0.50$ for all i, j. If randomized by transmission unit, then the assignments $\mathbf{Z}_1, \ldots, \mathbf{Z}_N$ could be made independently with $\Pr[\mathbf{Z}_i = 11] = \Pr[\mathbf{Z}_i = 01] = \Pr[\mathbf{Z}_i = 10] = \Pr[\mathbf{Z}_i = 00] = 0.25$ for all i. Under both assignment mechanisms in large samples, within each principal stratum each of the four possible $\mathbf{Z}_i$ would be assigned to one quarter of the transmission units. Other randomization probabilities would be possible.

3.1 Identifiability of $\mathrm{CVE}_I^{12}(0)$ and $\mathrm{CVE}_I^{21}(0)$

We consider first identifying $\mathrm{CVE}_I^{12}(0)$. The numerator of (12) is identifiable because transmission units assigned $\mathbf{Z}_i = (10)$ where $S_{i1}^{obs} = 1$ must be members of principal stratum 11. Therefore, under the assumptions above,

$$E[Y_{i2}(10) | \mathbf{S}_{i1} = (1111)] = E\left[Y_{i2}^{obs} | S_{i1}^{obs} = 1, \mathbf{Z}_i = (10)\right], \tag{27}$$

where, in large samples, $E\left[Y_{i2}^{obs} | S_{i1}^{obs} = 1, \mathbf{Z}_i = (10)\right] = \mathrm{SAR}_{10}^{12}$. In the following, even when not explicitly noted, we are assuming large samples.

Next we show that the denominator of (12), $E[Y_{i2}(00) | \mathbf{S}_{i1} = (1111)]$, is not identifiable without further assumptions. From Table 2, it can be seen that transmission units with $S_{i1}^{obs} = 1$, $\mathbf{Z}_i = (00)$ may be members of one of five principal strata, in particular strata 3, 5, 6, 8 and 11. Fortunately, assuming strongly ignorable vaccination assignment (26), the observable data do provide information about the proportion of such transmission units that can be expected to belong to the principal stratum 11 with $\mathbf{S}_{i1} = (1111)$ such that large-sample bounds can be derived (Section 3.4).

We show the proportion of units with $\mathbf{Z}_i = (00)$, $S_{i1}^{obs}=1$ in stratum 11 is identifiable. From assumption (26), it follows we can identify f from the transmission units with $\mathbf{Z}_i = (10)$, $S_{i1}^{obs}=1$ because these transmission units must be in principal stratum 11 with $\mathbf{S}_{i1} = (1111)$. Thus, in large samples

$$f=\frac{\sum_i I\left(\mathbf{Z}_i=(10), S_{i1}^{obs}=1\right)}{\sum_i I\left(\mathbf{Z}_i=(10)\right)}. \tag{28}$$

Similarly, in large samples,

$$b+c+d+e+f=\frac{\sum_i I\left(\mathbf{Z}_i=(00), S_{i1}^{obs}=1\right)}{\sum_i I\left(\mathbf{Z}_i=(00)\right)}. \tag{29}$$

By assumption (5), the denominators in (28) and (29) will be nonzero.

The proportion of transmission units in principal stratum 11 of those observed with $\mathbf{Z}_i = (00)$, $S_{i1}^{obs}=1$ from the five principal strata 3, 5, 6, 8, and 11 is identified from (28) and (29), giving

$$\delta=\frac{f}{b+c+d+e+f}=\frac{\sum_i I\left(\mathbf{Z}_i=(10), S_{i1}^{obs}=1\right)/\sum_i I\left(\mathbf{Z}_i=(10)\right)}{\sum_i I\left(\mathbf{Z}_i=(00), S_{i1}^{obs}=1\right)/\sum_i I\left(\mathbf{Z}_i=(00)\right)}, \tag{30}$$

Though we can identify the proportion of transmission units in the principal stratum of interest, we do not know which of the transmission units with $\mathbf{Z}_i = (00)$, $S_{i1}^{obs}=1$ they are. If the strata 3, 5, 6, 8 are empty, (and by symmetry, 2, 4, 7, and 9 are empty), then δ = 1. Analogous arguments hold for the identifiability of $\mathrm{CVE}_I^{21}(0)$ in (14).

3.2 Identifiability of $\mathrm{CVE}_I^{12}(1)$ and $\mathrm{CVE}_I^{21}(1)$

We now consider identifying $\mathrm{CVE}_I^{12}(1)$. The numerator of (13) is identifiable because, as seen in Table 2, transmission units with $\mathbf{Z} = (11)$, $S_{i1}^{obs}=1$ must be members of one of the three principal strata 8, 9, and 11, with $\mathbf{S}_{i1} \in \{(0101), (1101), (1111)\}$. Under the assumptions above,

$$E\left[Y_{i2}(11) \,|\, \mathrm{S}_{i1} \in \{(0101), (1101), (1111)\}\right] = E\left[Y_{i2}^{obs} | S_{i1}^{obs}=1, \mathbf{Z}_i=(11)\right], \tag{31}$$

where in large samples, $E\left[Y_{i2}^{obs} | S_{i1}^{obs}=1, \mathbf{Z}_i=(11)\right] = \mathrm{SAR}_{11}^{12}$.

Next we show that the denominator of (13), $E[Y_{I2}(01)|\mathbf{S}_{i1} \in \{(0101), (1101), (1111)\}]$, is not identifiable without further assumptions. From Table 2, it can be seen that transmission units with $S_{i1}^{obs}=1$ when assigned $\mathbf{Z}_i = (01)$ may be members of one of eight principal strata, in particular strata 3 through 9 and 11. We cannot identify which of the transmission units with $\mathbf{Z}_i = (01)$, $S_{i1}^{obs}=1$ are in the three principal strata 8, 9, and 11, of interest. However, we can identify the proportion of the transmission units with $\mathbf{Z}_i = (01)$, $S_{i1}^{obs}=1$ who are in the three principal strata of interest. In large samples,

$$d+e+f=\frac{\sum_i I\left(\mathbf{Z}_i=(11), S_{i1}^{obs}=1\right)}{\sum_i I\left(\mathbf{Z}_i=(11)\right)}. \quad (32)$$

Similarly,

$$b+2c+2d+2e+f=\frac{\sum_i I\left(\mathbf{Z}_i=(01), S_{i1}^{obs}=1\right)}{\sum_i I\left(\mathbf{Z}_i=(01)\right)}. \quad (33)$$

By assumption (5), the denominators in (32) and (33) will be nonzero. Therefore, the proportion of transmission units with $\mathbf{Z}_i = 01$, $S_{i1}^{obs}=1$ in the three principal strata of interest 8, 9, and 11 that make up the denominator of the causal estimand (13) is identified from (32) and (33):

$$\gamma=\frac{d+e+f}{b+2c+2d+2e+f}=\frac{\sum_i I\left(\mathbf{Z}_i=(11), S_{i1}^{obs}=1\right)/\sum_i I\left(\mathbf{Z}_i=(11)\right)}{\sum_i I\left(\mathbf{Z}_i=(01), S_{i1}^{obs}=1\right)/\sum_i I\left(\mathbf{Z}_i=(01)\right)}. \quad (34)$$

Though we can identify the proportion of transmission units in the three principal strata, we do not know which transmission units they are. When strata 3 through 9 are empty (and by symmetry stratum 2 is empty), $\gamma = 1$. Similar arguments hold for the identification of CVE_I^{21} in (15).

3.3 Identifiability of other estimands

The identifiability of $\mathrm{CVE}_I^c(0)$ and $\mathrm{CVE}_I^c(1)$ follow from the definitions of $\mathrm{CVE}_I^c(0)$ and $\mathrm{CVE}_I^c(1)$ in (17) and (18) and the arguments in Sections 3.1 and 3.2. The numerators in both causal estimands are identifiable from the observable data, but the denominators are not.

The VE_I^{net} and VE_S^{net} estimands in Section 2.5 are defined in terms of the expectations of the observed data. In large samples, the $E\left[Y_{ik}^{obs}|S_{ij}^{obs}=1, \mathbf{Z}_i=\left(Z_{ij}=r, Z_{ik}=s\right)\right]=\mathrm{SAR}_{rs}^{jk}$, and the $E\left[Y_i^{c,obs}|Y_{c,i}^{obs}=1, \mathbf{Z}_i=(r,s)\right]=\mathrm{SAR}_{rs}$, $r, s \in \{0, 1\}$, $j, k \in \{1, 2\}$, identifying all of the estimands. For example, equivalent to the standard method for estimating vaccine efficacy for infectiousness in (2),

$$\widehat{\mathrm{VE}}_I^{net,c}(0)=1-\frac{\mathrm{SAR}_{10}}{\mathrm{SAR}_{00}} \quad \text{and} \quad \widehat{\mathrm{VE}}_I^{net,c}(1)=1-\frac{\mathrm{SAR}_{11}}{\mathrm{SAR}_{01}}. \quad (35)$$

Under assumption (26), the causal protective effect of vaccination against infection from outside the transmission unit is identifiable. For example, from (26) it follows

$$\mathrm{VE}_{S,1}(0)=1-\frac{E\left[S_{i1}^{obs}|\mathbf{Z}_i=(10)\right]}{E\left[S_{i1}^{obs}|\mathbf{Z}_i=(00)\right]} \quad \text{and} \quad \mathrm{VE}_{S,1}(1)=1-\frac{E\left[S_{i1}^{obs}|\mathbf{Z}_i=(11)\right]}{E\left[S_{i1}^{obs}|\mathbf{Z}_i=(01)\right]}. \quad (36)$$

Analogous expressions hold for $\mathrm{VE}_{S,2}(0)$, $\mathrm{VE}_{S,2}(1)$, $\mathrm{VE}_{S,c}(0)$, and $\mathrm{VE}_{S,c}(1)$. From (30) and (34), δ and γ are identifiable and, in large samples, equal the ratios of the two expectations of the right and left equations in (36). Thus in large samples,

$$\widehat{\mathrm{VE}}_{S,1}(0)=\widehat{\mathrm{VE}}_{S,2}(0)=\widehat{\mathrm{VE}}_{S,c}(0)=1-\delta, \quad \text{and}$$
$$\widehat{\mathrm{VE}}_{S,1}(1)=\widehat{\mathrm{VE}}_{S,2}(1)=\widehat{\mathrm{VE}}_{S,c}(1)=1-\gamma.$$

When δ = 1, then $\widehat{\mathrm{VE}}_{S,k}(0)=0$, $k \in \{1, 2, c\}$ and the vaccine has no protective effect on infection from outside the transmission unit. Similarly, when γ = 1, then $\widehat{\mathrm{VE}}_{S,k}(1)=0$, $k \in \{1, 2, c\}$.

3.4 Large-sample bounds

We have shown that the proportion in sets of principal strata in Table 2 can be identified from the observed data under the given assumptions. We can use these proportions to derive large-sample bounds on the various CVE_I estimands similar to Zhang and Rubin (2003). Considering first $\mathrm{CVE}_I^{12}(0)$ (12), the proportion δ = $f/(b + c + d + e + f)$ of the transmission units with $\mathbf{Z}_i = (00)$, $S_{i1}^{obs}=1$ from the five principal strata 3, 5, 6, 8, and 11 in principal stratum 11 is identified. Depending on the observed data, we can determine the minimum and maximum number of these transmission units with $Y_{i2}^{obs}=1$ that could be in principal stratum 11, setting bounds on the causal $\mathrm{CVE}_I^{12}(0)$. The proof is in Appendix A. The results assume (26). The upper and lower bounds on the causal $\mathrm{CVE}_I^{12}(0)$ are

$$\widehat{\mathrm{CVE}}_I^{12,up}(0)=1-\frac{\mathrm{SAR}_{10}^{12}}{\mathrm{SAR}_{00}^{12}}\left\{\delta I\left[\delta>\mathrm{SAR}_{00}^{12}\right]+\mathrm{SAR}_{00}^{12}I\left[\delta\le\mathrm{SAR}_{00}^{12}\right]\right\} \tag{37}$$

$$\widehat{\mathrm{CVE}}_I^{12,low}(0)=\begin{cases}-\infty & \text{if } 1-\delta\ge\mathrm{SAR}_{00}^{12},\\ 1-\frac{\mathrm{SAR}_{10\delta}^{12}}{\mathrm{SAR}_{00}^{12}-(1-\delta)} & \text{if } 1-\delta<\mathrm{SAR}_{00}^{12}.\end{cases} \tag{38}$$

From (37) and (38), if δ = 1, then $\widehat{\mathrm{CVE}}_I^{12,up}(0)=\widehat{\mathrm{CVE}}_I^{12,low}(0)=\widehat{\mathrm{VE}}_I^{net,12}(0)$ The upper and lower bounds on the estimate of the causal $\mathrm{CVE}_I^{12}(1)$ are

$$\widehat{\mathrm{CVE}}_I^{12,up}(1)=1-\frac{\mathrm{SAR}_{11}^{12}}{\mathrm{SAR}_{01}^{12}}\left\{\gamma I\left[\gamma>\mathrm{SAR}_{01}^{12}\right]+\mathrm{SAR}_{01}^{12}I\left[\gamma\le\mathrm{SAR}_{01}^{12}\right]\right\} \tag{39}$$

$$\widehat{\mathrm{CVE}}_I^{12,low}(1)=\begin{cases}-\infty & \text{if } 1-\gamma\ge\mathrm{SAR}_{01}^{12},\\ 1-\frac{\mathrm{SAR}_{11\gamma}^{12}}{\mathrm{SAR}_{01}^{12}-(1-\gamma)} & \text{if } 1-\gamma<\mathrm{SAR}_{01}^{12}.\end{cases} \tag{40}$$

From (39) and (40), if γ = 1, then $\widehat{\mathrm{CVE}}_I^{12,up}(1)=\widehat{\mathrm{CVE}}_I^{12,low}(1)=\widehat{\mathrm{VE}}_I^{net,12}(1)$. The upper and lower bounds for the estimates of $\mathrm{CVE}_I^{21}(0)$ and $\mathrm{CVE}_I^{21}(1)$ are analogous.

If the labels $j = 1$ and $j = 2$ have been arbitrarily assigned, then

$$\widehat{\mathrm{CVE}}_I^{c,up}(0)=1-\frac{\mathrm{SAR}_{10}}{\mathrm{SAR}_{00}}\left\{\delta I\left[\delta>\mathrm{SAR}_{00}\right]+\mathrm{SAR}_{00}I\left[\delta\le\mathrm{SAR}_{00}\right]\right\} \tag{41}$$

$$\widehat{\mathrm{CVE}}_I^{c,low}(0) = \begin{cases} -\infty & \text{if } 1-\delta \geq \mathrm{SAR}_{00}, \\ 1 - \frac{\mathrm{SAR}_{10^\delta}}{\mathrm{SAR}_{00}-(1-\delta)} & \text{if } 1-\delta < \mathrm{SAR}_{00}. \end{cases} \quad (42)$$

From (41) and (42), if δ = 1, then $\widehat{\mathrm{CVE}}_I^{c,up}(0) = \widehat{\mathrm{CVE}}_I^{c,low}(0) = \widehat{\mathrm{VE}}_I^{net,c}(0)$. Also,

$$\widehat{\mathrm{CVE}}_I^{c,up}(1) = 1 - \frac{\mathrm{SAR}_{11}}{\mathrm{SAR}_{01}} \{\gamma I[\gamma > \mathrm{SAR}_{01}] + \mathrm{SAR}_{01} I[\gamma \leq \mathrm{SAR}_{01}]\} \quad (43)$$

$$\widehat{\mathrm{CVE}}_I^{c,low}(1) = \begin{cases} -\infty & \text{if } 1-\gamma \geq \mathrm{SAR}_{01}, \\ 1 - \frac{\mathrm{SAR}_{11}\gamma}{\mathrm{SAR}_{01}-(1-\gamma)} & \text{if } 1-\gamma < \mathrm{SAR}_{01}, \end{cases} \quad (44)$$

From (43) and (44), if $\boldsymbol{\gamma} = 1$, then $\widehat{\mathrm{CVE}}_I^{c,up}(1) = \widehat{\mathrm{CVE}}_I^{c,low}(1) = \widehat{\mathrm{VE}}_I^{net,c}(1)$. We conjecture these bounds are sharp, however a formal proof along the of Imai (2008) is beyond the scope of this paper.

3.5 Identification under assumption of no selection bias

The assumption of no selection bias is one selection model that identifies the causal estimands from the observed data. The strongest assumption of no selection bias is that the expected secondary transmission outcome is the same across all strata where there is a primary case for a particular $\mathbf{z}_i$, i.e., $E[Y_{ik}(\mathbf{z}_i) | S_{ij}(\mathbf{z}_i) = 1, \mathbf{S}_i] = E[Y_{ik}(\mathbf{z}_i) | S_{ij}(\mathbf{z}_i) = 1]$ for all $\mathbf{z}_j$. However, weaker assumptions of no selection bias are su cient to identify the causal estimands. We consider first the denominator of (12). Extrapolating from Hudgens and Halloran (2006), the weaker assumption of no selection bias with respect to (12) is

$$E[Y_{i2}(00) | \mathbf{S}_{i1} \in \{(1100), (1100), (1100), (1101)\}] = E[Y_{i2}(00) | \mathbf{S}_{i1} = (1111)] \quad (45)$$

Making this assumption for the four strata 3, 5, 6, and 8 identifies the denominator of (12) from the observed data, thus $E[Y_{i2}(00) | \mathbf{S}_{i1} = (1111)] = E\left[Y_{i2}^{obs} | S_{i1}^{obs} = 1, \mathbf{Z}_i = (00)\right]$. Then $\widehat{\mathrm{VE}}_I^{net,12}(0)$ is a consistent estimator of the causal $\mathrm{CVE}_I^{12}(0)$.

Considering the denominator of (13), the weaker assumption of no selection bias is that the $E[Y_{i2}(01) | S_{i1} \in \{(1100), (0100), (1100), (1100), (1100)\}]$ in the five principal strata 3 through 7 with $S_{i1}(01) = 1$ not in (13) equals the $E[Y_{i2}(01) | S_{i1} \in \{(1101), (0101), (1111)\}]$ in the three principal strata 8, 9, and 11 in (13). This assumption identifies the denominator of (13) from the observed data, and $E[Y_{i2}(01) | \mathbf{S}_{i1} \in \{(0101), (1101), (1111)\}] = E\left[Y_{i2}^{obs} | \mathbf{Z}_i = (01), S_{i1}^{obs} = 1\right]$. Then $\widehat{\mathrm{VE}}_I^{net,12}(1)$ is a consistent estimator of the causal $\mathrm{CVE}_I^{12}(1)$. Similar arguments hold for identifying the other causal CVE_I estimands under the assumption of no selection bias.

3.6 Example

Here we demonstrate identifiability and estimation of bounds of the causal estimands in Section 2.4 from data observable in a vaccine study. We continue the example begun in Table 1, a randomized, controlled vaccine study in $N = 11,000$ transmission units of size two. The assignments $Z_{11}, \ldots, Z_{N1}, Z_{12}, \ldots, Z_{N2}$ are independent with $\Pr[Z_{ij} = 1] = 0.50$ for all i, j. Within each principal stratum, each of the four possible $\mathbf{Z}_i$ are assigned to one quarter of the transmission units. We assume the labels 1 and 2 were arbitrarily assigned, and for brevity, we focus on the combined estimands $\mathrm{CVE}_I^c(0)$ and $\mathrm{CVE}_I^c(1)$. In the combined

estimands and estimators, the two people in the transmission unit are ordered such that $j = 1$ is the primary case and $j = 2$ is the exposed person if there is a primary case in the household. This design would also allow inference about the causal estimands (12)–(15) through the large sample bounds described in Section 3.4.

Table 4 contains the observed distribution of the transmission units by status of the primary case of Table 1, including also transmission units with no primary case. The data in Table 4 were generated with knowledge of the underlying principal strata (Appendix B, Table 8), but that knowledge is not necessary for the results here.

From Table 1 and (35), $\widehat{\mathrm{VE}}_I^{net,c}(0) = 0.50$ and $\widehat{\mathrm{VE}}_I^{net,c}(1) = 0.50$, equal to the traditional estimates in (4). Under the assumption of no selection bias, these would be consistent estimators of the causal estimands $\mathrm{CVE}_I^c(0)$ and $\mathrm{CVE}_I^c(1)$. However, this is an unverifiable assumption. From the data on $(S_{i1}^{obs}, S_{i2}^{obs}, \mathbf{Z}_i)$ in Table 4, the principal strata in Table 2, and the assumptions above, we can identify δ and γ for this example. Transmission units with $\mathbf{Z}_i = (00)$ and $\left(S_{i1}^{obs}, S_{i2}^{obs}\right) = (0, 0)$ must be in stratum 1, so the proportion $\alpha = 250/2750 = 1/11$. All other transmission units with $\mathbf{Z}_i = (00)$ have $\left(S_{i1}^{obs}, S_{i2}^{obs}\right) = (1, 0)$, so $2(b + c + d + e + f) = 2500/2750$. From either of these, $(b + c + d + e + f) = 5/11$ is identified. All of the 500 transmission units with $\mathbf{Z}_i = (01)$ or $\mathbf{Z}_i = (10)$ with a vaccinated primary case must be in stratum 10 or 11, so $f = 500/5500 = 1/11$. Thus $\delta = f/(b + c + d + e + f) = 0.2$.

Similarly, transmission units with $\mathbf{Z}_i = (11)$ and $\left(S_{i1}^{obs}, S_{i2}^{obs}\right) = (1, 0)$ must be in one of strata 6 through 11, thus $2(d + e + f) = 1500/2750$, and $(d + e + f) = 3/11$. Those transmission units with $\mathbf{Z}_i$ such that the primary case is unvaccinated and the exposed person is vaccinated must be in strata 2 through 11, whereby strata 4 through 9 are represented twice when collapsing over $\mathbf{Z}_i = (01)$ and $\mathbf{Z}_i = (10)$. Thus, $(b+2c+2d+2e+f) = 4000/5500 = 8/11$, and $\gamma = (d + e + f)/(b + 2c + 2d + 2e + f) = 3/8 = 0.375$. Thus we have illustrated that δ and γ can be identified from the observed data under the assumptions without knowledge of the underlying principal strata.

The bounds depend on the relation of the distributions of the basic principal strata and the $\mathbf{Y}_i$. From (41) and (42), the bounds on $\mathrm{CVE}_I^c(0)$ depend on the relation of SAR_{00} to δ, and from (43) and (44), the bounds on $\mathrm{CVE}_I^c(1)$ depend on the relation of SAR_{01} to δ. Based on Table 1, $\mathrm{SAR}_{00} = 0.90$, so $\delta < \mathrm{SAR}_{00}$, $(1 - \delta) < \mathrm{SAR}_{00}$, and $\mathrm{SAR}_{01} = 0.70$, so $\gamma < \mathrm{SAR}_{01}$ and $1 - \gamma < \mathrm{SAR}_{01}$. Thus,

$$\widehat{\mathrm{CVE}}_I^{c,up}(0) = 1 - \mathrm{SAR}_{10} = 1 - 0.45 = 0.55,$$

$$\widehat{\mathrm{CVE}}_I^{c,low}(0) = 1 - \frac{\mathrm{SAR}_{11}\delta}{\mathrm{SAR}_{01} - (1 - \delta)} = 1 - \frac{0.45 * 0.20}{0.90 - 0.80} = 0.1,$$

$$\widehat{\mathrm{CVE}}_I^{c,up}(1) = 1 - \mathrm{SAR}_{11} = 1 - 0.35 = 0.65,$$

$$\widehat{\mathrm{CVE}}_I^{c,low}(1) = 1 - \frac{\mathrm{SAR}_{11}\gamma}{\mathrm{SAR}_{01} - (1-\gamma)} = 1 - \frac{0.35 * 0.375}{0.70 * 0.675} = -4.25.$$

In summary, from (35), the two stratified net estimates of vaccine effect on infectiousness are $\widehat{\mathrm{VE}}_I^{net,c}(0) = 1 - 0.45/0.90 = 0.5$ and $\widehat{\mathrm{VE}}_I^{net,c}(1) = 1 - 0.35/0.70 = 0.5$. Under the assumption of no selection bias, $\widehat{\mathrm{VE}}_I^{net,c}(z)$ is a consistent estimator of $\mathrm{CVE}_I^c(1)$. Without this assumption, the large sample bounds provide a range of estimates of $\mathrm{CVE}_I^c(z)$. For these data the bounds on $\mathrm{CVE}_I^c(1)$ are wide (−4.25, 0.65), however the bounds on $\mathrm{CVE}_I^c(0)$ are narrower (0.1, 0.55) and are informative in the sense that the null value of no vaccine effect on infectiousness is excluded.

3.7 Non-random labels

If the labels 1 and 2 are not randomly assigned, but represent a covariate, say 1 was the younger sibling and 2 was the older sibling, then we might be interested in the causal estimands $\mathrm{CVE}_I^{12}(0)$ and $\mathrm{CVE}_I^{12}(1)$ separately from $\mathrm{CVE}_I^{21}(0)$ and $\mathrm{CVE}_I^{21}(1)$ rather than the composite causal estimands. In Table 2, we could no longer assume that the proportions of the population in the symmetric pairs of principal strata, such as (2,3), (4,5) and so forth, were equal. We would need to allow for proportions b_1, b_2, c_1, c_2, d_1, d_2, e_1, e_2, f_1, and f_2. Identifiability arguments analogous to those in Sections 3.1 and 3.2 would still hold with the expanded notation, and yield different proportions δ_1, δ_2, γ_1 and γ_2 for estimating the different bounds on the two pairs of causal estimands.

4 The Simple Scenario

We now consider the simplest scenario by making two key assumptions. This simple scenario is the most straightforward extension of Hudgens and Halloran (2006) to the situation where the postinfection outcome is secondary transmission to another individual. First, we assume individuals within the transmission unit can be ordered such that only the first individual ($j = 1$) is exposed to infection from outside the transmission unit. This assumption implies individual $j = 1$ can never be the secondary case and individual $j = 2$ can never be the primary case. Formally,

$$\text{Unique primary case:} \quad Y_{i1}(\mathbf{z}_i) = S_{i2}(\mathbf{z}_i) = 0 \quad \text{for all} \quad i, \mathbf{z}_i. \tag{46}$$

The second key assumption is that only individual $j = 1$ can be assigned to vaccine or control, and individual $j = 2$ is not vaccinated, assumed equivalent to control:

$$\text{Unique assignment:} \quad \mathbf{Z}_i = (Z_{i1}, 0), \quad Z_{i1} \in \{0, 1\} \quad \text{for all} \quad i. \tag{47}$$

Under assumption (47) individual $j = 2$ is never vaccinated, such that the potential outcomes depend only on the vaccination assignment of individual $j = 1$, and thus can be written more simply as $Y_{ij}(z_{i1})$ and $S_{ij}(z_{i1})$. However, we maintain the notation that explicitly shows individual $j = 2$ is never vaccinated, so we write $Y_{ij}(z_{i1}, 0)$ and $S_{ij}(z_{i1}, 0)$. Under assumptions (46) and (47), we define the basic principal stratification according to the joint potential infection outcomes $\mathbf{S}_{i1} = (S_{i1}(00), S_{i1}(10))$ in individual $j = 1$ only. There are just four possible combinations of $(S_{i1}(00), S_{i1}(10))$. We make the usual monotonicity assumption that the protective effect of vaccination is nonnegative. In this scenario, this is a special case of assumption (7) with $j = 1$, $(j) = 2$, and $z = 0$. Thus, there are three feasible basic principal strata (Table 5), called immune ($\mathbf{S}_{i1} = (00)$), protected ($\mathbf{S}_{i1} = (10)$), and doomed ($\mathbf{S}_{i1} = (11)$).

We assume as in (26) that vaccination assignment is independent of the potential outcomes. Formally,

$$\text{Independence:} \quad \mathbf{Z}_i = (Z_{i1}, 0) \perp \{\mathbf{S}_i, \mathbf{Y}_i\} \quad \text{for all} \quad i. \tag{48}$$

As an example of this simple scenario the transmission unit could be defined as a man and woman engaged in a monogamous sexual relationship with only the man exposed to HIV outside the partnership through injecting drug use (Longini et al. 1999). A vaccine trial might randomize only the injecting drug users to vaccine or control (Datta et al. 1998).

4.1 Estimands

In Table 5, the doomed basic principal stratum is the only stratum in which both potential postinfection transmission endpoints, $\mathbf{Y}_i = (Y_{i2}(00), Y_{i2}(10))$, and thus their joint distribution, are defined. In other words, the causal effect of the vaccine on infectiousness must be defined by contrasts between $[Y_{i2}(00)|\mathbf{S}_{i1} = (11)]$ and $[Y_{i2}(10)|\mathbf{S}_{i1} = 11]$. Analogous to (12), we define the causal estimand

$$\text{CVE}_I^{12}(0) = 1 - \frac{E[Y_{i2}(10)|\mathbf{S}_{i1} = (11)]}{E[Y_{i2}(00)|\mathbf{S}_{i1} = (11)]}, \tag{49}$$

i.e., the relative reduction in the probability of secondary transmission due to vaccination in the principal stratum of transmission units where individual $j = 1$ becomes infected regardless of vaccination status.

Under (46) and (47), the estimand for the net vaccine effect on infectiousness is

$$\text{VE}_I^{net,12}(0) = 1 - \frac{E\left[Y_{i2}^{obs}|S_{i1}^{obs} = 1, \mathbf{Z}_i = (10)\right]}{E\left[Y_{i2}^{obs}|S_{i1}^{obs} = 1, \mathbf{Z}_i = (00)\right]} = 1 - \frac{E[Y_{i2}(10)|S_{i1}(10) = 1]}{E[Y_{i2}(00)|S_{i1}(00) = 1]},$$

with the second equality following from (48). Again, in general, $\text{VE}_I^{net,12}(0)$ does not have a causal interpretation because the set of individuals with $S_{i1}(10) = 1$ is not necessarily identical to the set of individuals with $S_{i1}(00) = 1$.

Under assumptions (46) and (47), we can define another estimand for the effect of vaccination on the outcome Y in individual $j = 2$ that does not condition on the transmission unit having a primary case:

$$\text{VE}^{ITT}(0) = 1 - \frac{E[Y_{i2}(10)S_{i1}(10)]}{E[Y_{i2}(00)S_{i1}(00)]}, \tag{50}$$

where we adopt the convention $Y_{i2}(\mathbf{z}_i)S_{i1}(\mathbf{z}_i) = 0$ when $S_{i1}(\mathbf{z}_i) = 0$ and $Y_{i2}(\mathbf{z}_i) = *$, $\mathbf{z}_i \in \{(10), (00)\}$. We consider such an estimand intent-to-treat (ITT) because it does not condition on the post-treatment variable S_{i1}^{obs}, i.e., it incorporates all individuals $j = 2$ according to the vaccination assignment of the individual $j = 1$ in their transmission unit. $\text{VE}^{ITT}(0)$ does have a causal interpretation. Estimand (50) is not an estimand for the vaccine effect on infectiousness, but rather it captures an indirect effect of vaccination (Halloran and Struchiner 1991) because it describes the indirect effect of treatment (vaccine) on an individual of the treatment received in others in the same group where interference can occur. The indirect effect does not condition on the infection status of the others in the group. The $\text{VE}^{ITT}(0)$ estimand is a special case of the population average indirect effect causal estimand defined by Hudgens and Halloran (2008). In the scenario in this paper, the

population of groups is the population of transmission units of size two. This type of study is called a mini-community design (Halloran et al. 2010). Examples include a study of indirect effects of pneumococcal vaccination on unvaccinated household members (Millar et al. 2008) and a study of indirect effects of pertussis vaccination on unvaccinated siblings (Trollfors et al. 1998).

Similar to (25), we define the causal vaccine effect on infection by

$$\mathrm{VE}_{s,1}(0) = -\frac{E[S_{i1}(10)]}{E[S_{i1}(00)]}, \tag{51}$$

i.e., the relative average causal effect of vaccination on infection in individual $j = 1$ exposed outside the transmission unit when $j = 2$ is not vaccinated.

4.2 Identifiability and large-sample bounds

The causal $\mathrm{CVE}_{I}^{12}(0)$ in (49) is not identifiable without further assumptions. The numerator of (49) is identifiable because transmission units with $S_{i1}^{obs}=1$, $\mathbf{Z}_i = (10)$ must be members of the $\mathbf{S}_{i1} = (11)$ doomed principal stratum. Under the assumptions made above, then

$$E[Y_{i2}(10)|S_{i1}(11)] = E\left[Y_{i2}^{obs}|S_{i1}^{obs}=1, \mathbf{Z}_i=(10)\right]. \tag{52}$$

The denominator of (49), $E[Y_{i2}(00)|\mathbf{S}_{i1} = (11)]$, is not identifiable without further assumptions, because from Table 5 it can be seen that transmission units with $S_{i1}^{obs}=1$, $\mathbf{Z}_i = (00)$ may be members of either the protected or doomed principal stratum. Assuming independent vaccination assignment (48), the observable data provide information about the proportion of such transmission units that can be expected to belong to the $\mathbf{S}_{i1} = (11)$ doomed principal stratum, such that large-sample bounds can be derived. Using arguments as in Section 3, the proportion f in the doomed principal strata and the proportion $d + f$ in the combined protected and doomed strata are identifiable. Thus, the proportion of the transmission units with S_{i1}^{obs}, $\mathbf{Z}_i = (00)$ in the doomed principal stratum, say $\rho = f/(d + f)$, is identifiable. The upper and lower bounds for the causal $\mathrm{CVE}_{I}^{12}(0)$ estimand (49) are similar to those in (37) and (38) with ρ replacing δ. These bounds are analogous to (23) and (26) in Hudgens and Halloran (2006), with $\widehat{\mathrm{VE}}_S$ in that paper replaced by $1 - \rho$ here and other appropriate adjustments in notation.

The estimands $\mathrm{VE}_{I}^{net,12}(0)$, $\mathrm{VE}^{ITT} = (0)$, and $\mathrm{VE}_{S,1}(0)$ are identifiable without further assumptions. Under assumption (48), it follows

$$\mathrm{VE}_{S,1}(0) = 1 - \frac{E\left[S_{i1}^{obs}|\mathbf{Z}_i=(10)\right]}{E\left[S_{i1}^{obs}|\mathbf{Z}_i=(00)\right]} \text{ and } \mathrm{VE}^{IIT}(0) = 1 - \frac{E\left[Y_{i2}^{obs}S_{i1}^{obs}|\mathbf{Z}_i=(10)\right]}{E\left[Y_{i2}^{obs}S_{i1}^{obs}|\mathbf{Z}_i=(00)\right]},$$

where we adopt the convention that $Y_{i2}^{obs}S_{i1}^{obs}=0$ when $S_{i1}^{obs}=0$ and $Y_{i2}^{obs}=*$. Define the attack rate $\mathrm{AR}_{1,(rs)}$ as the observed proportion of transmission units with vaccination status $\mathbf{Z}_i = (rs)$ with $S_{i1}^{obs}=1$. Define the attack rate $\mathrm{AR}_{2,(rs)}$ as the observed proportion of transmission units with vaccination status $\mathbf{Z}_i = (rs)$ with $Y_{i2}^{obs}=1$. Formally, for $r, s \in \{0, 1\}$,

$$\mathrm{AR}_{1,(rs)}=\frac{\Sigma_i I\left(\mathbf{Z}_i=(rs)\,,\,S_{i1}^{obs}=1\right)}{\Sigma_i I\left(\mathbf{Z}_i=(rs)\right)},\ \mathrm{AR}_{2,(rs)}=\frac{\Sigma_i I\left(\mathbf{Z}_i=(rs)\,,\,Y_{i2}^{obs}=1\right)}{\Sigma_i I\left(\mathbf{Z}_i=(rs)\right)}. \tag{53}$$

In large samples, $E\left[S_1^{obs}|\mathbf{Z}_i=(rs)\right]=\mathrm{AR}_{1,(rs)}$ and $E\left[Y_2^{obs}S_{1i}^{obs}|\mathbf{Z}_i=(rs)\right]=\mathrm{AR}_{2,(rs)}$. Thus,

$$\begin{aligned}\widehat{\mathrm{VE}}_I^{net,12}(0) &= 1-\frac{\mathrm{SAR}_{10}}{\mathrm{SAR}_{00}},\quad \widehat{\mathrm{VE}}_{S,1}(0)=1-\frac{\mathrm{AR}_{1,(10)}}{\mathrm{AR}_{1,(00)}}=1-\rho,\\ \widehat{\mathrm{VE}}^{IIT}(0) &= 1-\frac{\mathrm{AR}_{2,(10)}}{\mathrm{AR}_{2,(00)}}.\end{aligned} \tag{54}$$

The assumption of no selection bias is $E\left[Y_{i2}(00)\,|\mathbf{S}_{i1}=(11)\right]=E\left[Y_{i2}^{obs}|S_{i1}^{obs}=1,\mathbf{Z}_i=(00)\right]$. $\widehat{\mathrm{VE}}_I^{net,12}(0)$, is a consistent estimator of $\mathrm{CVE}_I^{12}(0)$ under the assumption of no selection bias. When ρ = 1, the protected principal stratum in Table 5 is empty, $\widehat{\mathrm{VE}}_{S,1}(0)=0$, and $\widehat{\mathrm{VE}}_I^{net,12}(0)$ is a consistent estimator of $\mathrm{CVE}_I^{12}(0)$.

5 Both Assigned Treatment, One Exposed Outside

Next we assume as in (5) of the general scenario (Section 2) both individuals have a non-zero probability of being assigned to vaccine or control, but as in (46) only individual $j = 1$ is exposed to infection outside the transmission unit. An extension of the HIV vaccine trial described in Section 4 in which the monogamous sexual partners are assigned to vaccine or control would be an example of this scenario. As in the simple scenario in Section 4, under the modified monotonicity assumption (7), there are three basic principal strata defined by the joint potential infection outcomes of individual $j = 1$ under vaccine and control. Interference does not affect the potential outcomes of individual $j = 1$ since we are assuming (46), i.e., only individual $j = 1$ is exposed from outside the transmission unit. There are now four possible postinfection strata in the protected basic principal stratum, and 16 in the doomed basic principal stratum. However, making the monotonicity assumption (11), the possibilities are reduced to three and nine (Table 6). We further assume that vaccine assignment is independent of their potential outcomes as in (26).

We now have two causal VE_I estimands analogous to $\mathrm{CVE}_I^{12}(0)$ in (12) and $\mathrm{CVE}_I^{12}(1)$ in (13) whereby now $\mathrm{CVE}_I^{12}(1)$ is defined in just the doomed principal stratum:

$$\mathrm{CVE}_I^{12}(1)=1-\frac{E\left[Y_{i2}(11)\,|\mathbf{S}_{i1}=(1111)\right]}{E\left[Y_{i2}(01)\,|\mathbf{S}_{i1}=(1111)\right]}. \tag{55}$$

Identifiability arguments proceed as previously. The large-sample bounds on the estimates of $\mathrm{CVE}_I^{12}(0)$ and $\mathrm{CVE}_I^{12}(1)$ are as in (37)–(38) and (39)–(40) with both δ and γ replaced by $\rho = f/(d + f)$, as in Section 4.

In this scenario, we can define the further estimands $\mathrm{VE}_I^{net,12}(0)$, $\mathrm{VE}_I^{net,12}(1)$, $\mathrm{VE}_S^{net,12}(0)$, $\mathrm{VE}_S^{net,12}(1)$, and $\mathrm{VE}_{S,1}(0)$ as in Section 2.5, and estimate the corresponding $\widehat{\mathrm{VE}}_I^{net,12}(0)$, $\widehat{\mathrm{VE}}_I^{net,12}(1)$, $\widehat{\mathrm{VE}}_S^{net,12}(0)$, $\widehat{\mathrm{VE}}_S^{net,12}(1)$, and $\widehat{\mathrm{VE}}_{S,1}(0)$. We can now define two ITT estimands, one in which $j = 2$ is not vaccinated, $\mathrm{VE}^{ITT}(0)$ as in (50), and one in which $j = 2$ is vaccinated, $\mathrm{VE}^{ITT}(1)$. Both are again special cases of the population average indirect effect causal estimand. The two ITT estimands are identifiable from the data as $\widehat{\mathrm{VE}}^{ITT}(0)=1-\mathrm{AR}_{2,(10)}/\mathrm{AR}_{2,(00)}$ and $\widehat{\mathrm{VE}}^{ITT}(1)=1-\mathrm{AR}_{2,(11)}/\mathrm{AR}_{2,(01)}$.

6 Non-Unique Primary Cases, One Assigned Vaccine

Next we assume as in Section 2, either individual may be the primary case, and as in (47), only individual $j = 1$ may be assigned vaccine or control. In this scenario, the vector $\mathbf{S}_i = (S_{i1}(00), S_{i1}(10), S_{i2}(00), S_{i2}(10))$ can take on $2^4 = 16$ possible values. We make three realistic exclusion restriction assumptions similar to those previously which decrease the number of $\mathbf{S}_i$ that are feasible. First, we assume as in (6) there is a unique primary case and no co-primary cases. This assumption implies $\sum_{j=1}^{2} S_{ij}(z_{i1}0) \leq 1$. Second, we assume as in (7) if $S_{i1}(10) = 1$, then $S_{i1}(00) = 1$. That is, if individual $j = 1$ becomes the primary case when vaccinated, s/he will also become the primary case when not vaccinated. Third, similar to (8), we assume if $S_{i2}(00) = 1$, then $S_{i2}(10) = 1$. That is, if individual $j = 2$ becomes the primary case when individual $j = 1$ is not vaccinated, then individual $j = 2$ will also become the primary case when individual $j = 1$ is vaccinated. Under these three exclusion restrictions, only six values of $\mathbf{S}_i$ are feasible, shown in Table 7. The right side of Table 7 displays the possible values of the potential secondary transmission outcomes $Y_{ij}(z_{i1}0)$ for each of the six feasible values of $\mathbf{S}_i$.

There are two basic principal strata within which causal estimands regarding secondary transmission to the same individual are well-defined, $\{i : \mathbf{S}_i = (1100)\}$ and $\{i : \mathbf{S}_i = (0011)\}$. Under assumption (47), only individual $j = 1$ may be vaccinated, implying only the former principal stratum provides information about the vaccine effect on infectiousness. Because $(S_{i1}(00), S_{i1}(10)) = (11)$ implies $\mathbf{S}_i = (1100)$ under the assumed exclusion restrictions, the causal estimand of interest is again $\mathrm{CVE}_I^{12}(0)$ as in the simplest scenario in (49). Identifiability and large-sample bounds arguments proceed as previously. We assume independent vaccination assignment (48). Because $S_{i1}^{obs} = 1$ given $\mathbf{Z}_i = (01)$ implies membership in the $\{i : \mathbf{S}_i = (1100)\}$ basic principal stratum, one can identify the numerator of $\mathrm{CVE}_I^{12}(0)$ and the proportion f in Table 7. However $S_{i1}^{obs} = 1$ given $\mathbf{Z}_i = (00)$ implies only that membership must be in one of three basic principal strata, namely $\{i : \mathbf{S}_i = (1100)\}$, $\{i : \mathbf{S}_i = (1001)\}$, or $\{i : \mathbf{S}_i = (1000)\}$. That is, the denominator of $\mathrm{CVE}_I^{12}(0)$ is not identifiable, but the proportion $d+e+f$ is. We define $\theta = f/(c+d+f)$. The bounds on $\mathrm{CVE}_I^{12}(0)$ are given in (37) and (38) with δ replaced by θ. Under the assumption of no selection bias, $\widehat{\mathrm{VE}}_I^{net,12}(0)$ is a consistent estimator of $\mathrm{CVE}_I^{12}(0)$. One can also estimate $\widehat{\mathrm{VE}}_S^{net,21}(0)$ in this scenario.

7 Discussion

In this paper, we have defined causal estimands for vaccine efficacy for infectiousness for four different scenarios of transmission units of size two. The causal estimands are defined within principal strata or unions of principal strata determined by the joint potential infection outcomes indicating who is a primary case under the four possible allocations of vaccine and control in the transmission unit. A series of exclusion restriction assumptions for each of the four scenarios enabled restriction of the number of feasible principal strata. Identifiability of the causal estimands for vaccine efficacy for infectiousness under the assumption that vaccine assignment was independent of the potential outcomes was considered and bounds for each estimand derived. An example demonstrated the use of the methods to analyze data that would be available in a study such as the general scenario.

Although our focus here was on causal estimands for vaccine efficacy for infectiousness, several other estimands were defined. Two different levels of protective effects of vaccination have been defined, one for protective effects of vaccination when exposed

outside the transmission unit, and one for protective effects of vaccination when exposed within the transmission unit. It is not uncommon that vaccination studies present two analyses, one that does not condition on known exposure to infection and one that does (see Kendrick and Eldering (1939) for an early example). However, their relation has never been studied in a causal inference setting and is deserving of further research. We defined two intent-to-treat causal estimands, $\mathrm{VE}^{ITT}(0)$ and $\mathrm{VE}^{ITT}(1)$ that are special cases of causal estimands for indirect effects in the setting of causal inference in the presence of interference (Halloran and Hudgens 2008).

The large sample bounds on the causal vaccine efficacy for infectiousness estimands assume extreme selection models. Future research could develop sensitivity analyses that explore potentially more realistic selection models. For the simple scenario in Section 4, the three approaches to sensitivity analysis for binary postinfection outcomes in Hudgens and Halloran (2006) immediately apply.

VanderWeele and Tchetgen (2011) also proposed causal estimands of vaccine efficacy for infectiousness. Considering the simplest case in Section 4, and stated in terms of our development, they derive a lower bound on the causal vaccine efficacy for infectiousness by assuming the probability of secondary transmission from unvaccinated primary cases in the protected stratum is not greater than that in the doomed principal stratum. The intuition behind this assumption is that people in the protected principal stratum are likely healthier than those in the doomed principal stratum, and thus less apt to transmit the infection to others. Thus, the observed $\widehat{\mathrm{VE}}_{l}^{net}(0)$ provides a lower bound on the causal $\mathrm{CVE}^{12}(0)$ in (49). However, this improvement in the lower bound is based on an assumption which is not verifiable from the observable data.

Several other avenues of future research could be pursued. In this paper we have assumed there is no interference across transmission units. This assumption might hold, if as stated in Section 2.2, the transmission units are transmission dynamically separated, geographically and/or socially. However, if there is interference between individuals from different transmission units, then one could re-define the transmission units to include all individuals between whom there is interference. For instance, if there is interference between individuals in different households that are geographically adjacent, then the transmission unit can be defined as a collection of adjacent households rather than as the individual households. For example, Préziosi and Halloran (2003) define as the transmission unit the compound, a collection of several small houses containing an extended family, to estimate vaccine efficacy for infectiousness. In that study, the number of exposed susceptible individuals ranged from 1 to 32 (interquartile range 2–8). For such a setting, the methods of this paper would need to be extended to allow for transmission units to have more than two individuals. Another approach to estimating vaccine efficacy based on the secondary attack rate models the interference among people across transmission units (Halloran et al. 2010, Chapter 11). Future research could develop methods for causal inference for this setting.

Throughout we have assumed that vaccine assignment was independent of potential outcomes. One could extend the methods in this paper to partially randomized or wholly observational studies. In the absence of randomized vaccination assignment, various methods could be adapted to address possible violations of the ignorable treatment assignment assumption. For example, one could relax this assumption by instead assuming independent vaccination assignment conditional on some set of baseline covariates (VanderWeele and Tchetgen 2011). Methods for sensitivity analysis to violation of the assumption of independent assignment mechanism could also be developed. For instance, following Rosenbaum (2010, Chapter 3), one could explore explore how inference about vaccine efficacy for infectiousness changes according to the strength of association assumed

between treatment assignment and some unmeasured confounder. Another issue in observational studies could be that transmission units would be ascertained only if they have a primary case. In this situation, the proportion of transmission units under each vaccine assignment with no primary cases would not be known. Further research is needed to draw inference about vaccine effects for infectiousness for this scenario.

Future research could also consider in more detail the consequences of the different exclusion restriction assumptions made in the different scenarios. We have not used information on $\mathbf{S}_i$ to inform exclusion restriction assumptions on $\mathbf{Y}_i$. Such exclusion restriction assumptions and their potential consequences for assumptions about selection bias need further research.

Here we have begun to set estimation of the effects of vaccination on reducing infectiousness of one person for another on the sound footing of causal inference. Much research remains to be done on this topic of considerable public health importance.

A Proof of Upper and Lower Bounds

Here we derive the upper and lower bounds for the denominator of the estimator of the causal estimand $\mathrm{CVE}_I^{12}(0)$ in (12). In Section 3.1, we show

$$\Pr[\mathbf{S}_{i1}=(1111)\,|S_{i1}(00)=1]=\frac{f}{b+c+d+e+f}=\delta$$

is identifiable. Next, note

$$\begin{aligned}
E\left[Y_{i2}^{obs}|\mathbf{Z}_i=(00),S_{i1}^{obs}=1\right] &= E[Y_{i2}(00)|\mathbf{Z}_i=(00),S_{i1}(00)]=1\\
&= \Pr[Y_{i2}(00)=1|S_{i1}(00)=1]\\
&= \Pr[Y_{i2}(00)=1|\mathbf{S}_{i1}=(1111)\,|S_{i1}(00)=1]\\
&+ \Pr[Y_{i2}(00)=1,\mathbf{S}_{i1}\neq(1111)\,|S_{i1}(00)=1]\\
&= \Pr[Y_{i2}(00)=1|\mathbf{S}_{i1}=(1111),S_{i1}(00)=1]\\
&\quad \Pr[\mathbf{S}_{i1}=(1111)\,|S_{i1}(00)=1]\\
&+ \Pr[Y_{i2}(00)=1|\mathbf{S}_{i1}\neq(1111),S_{i1}(00)=1]\\
&\quad \Pr[\mathbf{S}_{i1}\neq(1111)\,|S_{i1}(00)=1]\\
&= \Pr[Y_{i2}(00)=1|\mathbf{S}_{i1}=(1111)]\\
&\quad \Pr[\mathbf{S}_{i1}=(1111)\,|S_{i1}(00)=1]\\
&+ \Pr[Y_{i2}(00)=1|\mathbf{S}_{i1}\neq(1111)]\\
&\quad \Pr[\mathbf{S}_{i1}\neq(1111)\,|S_{i1}(00)=1]\\
&= \Pr[Y_{i2}(00)=1|\mathbf{S}_{i1}=(1111)]\,\delta\\
&+ \Pr[Y_{i2}(00)=1|\mathbf{S}_{i1}\neq(1111)]\,(1-\delta).
\end{aligned}$$

where the first and second equalities follow from the independence assumption (26) and $Y_{i2}(00)$ being binary. Therefore,

$$\Pr\left[Y_{i2}^{obs}=1|\mathbf{Z}_i=(00),S_{i1}^{obs}=1\right]=\Pr[Y_{i2}(00)=1|\mathbf{S}_{i1}=(1111)]\,\delta+\Pr[Y_{i2}(00)=1|\mathbf{S}_{i1}\neq(1111)]\,(1-\delta),\quad(56)$$

where $Pr\left[Y_{i2}^{obs}=1|\mathbf{Z}_i(00),S_{i1}^{obs}=1\right]$, δ, and 1 − δ are identifiable, and Pr[$Y_{i2}(00) = 1|\mathbf{S}_{i1} = (1111)$] is the estimand of interest. Following Hudgens and Halloran (2006), the upper bound on Pr[$Y_{i2}(00) = 1|\mathbf{S}_{i1} = (1111)$] would be achieved if either

$$\Pr[Y_{i2}(00)=1|\mathbf{S}_{i1}=(1111)]=1 \quad \text{or} \quad \Pr[Y_{i2}(00)=1|\mathbf{S}_{i1}\neq(1111)]=0.$$

If $\delta < \Pr[Y_{i2}(00)=1|\mathbf{Z}_i=(00), S_{i1}^{obs}=1]$, then $\Pr[Y_{i2}(00)=1|\mathbf{S}_{i1}=(1111)]=1$. Otherwise, $\Pr[Y_{i2}(00)=1|\mathbf{S}_{i1}\neq(1111)]=0$, which implies

$$\Pr[Y_{i2}(00)=1|\mathbf{S}_{i1}=(1111)]=\Pr\left[Y_{i2}^{obs}=1|\mathbf{Z}_i=(00), S_{i1}^{obs}=1\right]/\delta.$$

Thus, the upper bound on $\Pr[Y_{i2}(00)=1|\mathbf{S}_{i1}=(1111)]$ equals

$$\min\left\{1, \quad \frac{\Pr\left[Y_{i2}^{obs}=1|\mathbf{Z}_i=(00), S_{i1}^{obs}=1\right]}{\delta}\right\} \tag{57}$$

Similarly, the lower bound on $\Pr[Y_{i2}(00)=1|\mathbf{S}_{i1}=(1111)]$ equals

$$\max\left\{0, \quad \frac{\Pr\left[Y_{i2}^{obs}=1|\mathbf{Z}_i=(00), S_{i1}^{obs}=1\right]-(1-\delta)}{\delta}\right\} \tag{58}$$

Thus, the upper and lower bounds of $\mathrm{CVE}_I^{12}(0)$ are (37) and (38). Proofs of the upper and lower bounds of the other causal CVE_I estimands proceed similarly.

B Relation of Principal Strata to Usually Observed Data

Here we develop an example to illustrate the relation between the principal strata in Table 2 and the observed primary cases. Table 8 contains an example of the observed primary cases given the underlying principal strata. The study population has $N = 11,000$ transmission units of size two assumed to be equally distributed among the 11 principal strata, with 1000 transmission units in each. The labels 1 and 2 are assumed arbitrarily assigned. The assumption of equally distributed principal strata is not necessary for the results to hold. We have assumed the $Z_{11}, \ldots, Z_{N1}, Z_{12}, \ldots, Z_{N2}$ are independent with $\Pr[Z_{ij}=1]=0.50$ for all i, j. Thus, in large samples, in each principal stratum, one quarter of the transmission units would be assigned each of the four possible $\mathbf{Z}_i$ to vaccine or control. For each transmission unit i assigned $\mathbf{Z}_i$, whether there will be a primary case and which it will be is determined by the vector of potential outcomes $\mathbf{S}_i$ in Table 2.

For example, in principal stratum 2, in the 250 transmission units with $\mathbf{Z}_i=(00)$, individual $j=2$ is the primary case. In principal stratum 9, in the 250 transmission units with $\mathbf{Z}_i=(01)$, individual $j=1$ is the primary case. In general, we cannot determine which transmission units are in which principal stratum. However, the boldfaced numbers are those where it can be determined. In particular, the 250 transmission units with $\mathbf{Z}_i=(00)$ in which no one gets infected are known to be in stratum 1 under the assumptions. The 250 transmission units with $\mathbf{Z}_i=(01)$ in which the vaccinated individual $j=2$ is the primary case are in stratum 10. Similarly, the 250 transmission units with $\mathbf{Z}_i=(10)$ in which the vaccinated individual $j=1$ is the primary case are in stratum 11.

In the lower part of Table 8, the observed data on the primary cases are presented, first, in the situation that we distinguish individual $j=1$ and individual $j=2$, then below, if we did not distinguish them. For example, in the former case, in transmission units with $\mathbf{Z}_i=(01)$, there are 500 with no primary case, 2000 with an unvaccinated primary case, and 250 with a

vaccinated primary case. If the two individuals are not distinguished and one person in the transmission unit is vaccinated, and the other not, the central two columns are collapsed. In this case, with discordantly randomized transmission units, 1000 transmission units do not have a primary case, 4000 have an unvaccinated primary case, and 500 have a vaccinated primary case. Altogether, 2500 of the transmission units have no primary case. Of those that have a primary case, 2500 + 4000 = 6500 have an unvaccinated primary case and 500 + 1500 = 2000 have a vaccinated primary case. The data at the bottom of Table 8 are presented in Table 4 in the main text. The values of $\delta = f/(b + c + d + e + f) = 0.20$ and $\gamma = (d + e + f)/(b + 2c + 2d + 2e + f) = 3/8 = 0.375$ are easily computed from Table 8 when we know who is in each basic principal stratum. However, as demonstrated in the main text, δ and γ can be computed from the observed data under the assumptions.

Acknowledgments

This research was supported by National Institute of Allergy and Infectious Disease grants R01-AI085073 and R37-AI032042.

References

Becker N, Britton T, O'Neill P. Estimating vaccine effects from studies of outbreaks in household pairs. Stat Med. 2006; 25:1079–1093. [PubMed: 16287206]

Black H, Shinefeld B, Fireman S, et al. Efficacy, safety and immunogenicity of heptavalent pneumococcal conjugate vaccine in children. Northern California Kaiser Permanente Vaccine Study Center Group. Pediatr Infect Dis J. 2000; 19:187–195. [PubMed: 10749457]

Cisse B, Aaby P, Simondon F, Samb B, Soumaré M, Whittle H. Role of schools in the transmission of measles in rural Senegal: Implications for measles control in developing countries. Am J Epidemiol. 1999; 149:295–301. [PubMed: 10025469]

Clemens, Sack DA, Harris J, et al. J. Field trial of oral cholera vaccines. Lancet. 1986; 2:124–7. [PubMed: 2873397]

Cohen MS, Chen YQ, McCauley M, et al. Prevention of HIV-1 infection with early antiretroviral therapy. N Engl J Med. Jul 18.2011 10.1056/NEJMoa1105243, published online at NEJM.org.

Cox, D. Planning of Experiments. John Wiley and Sons, Inc; New York: 1958.

Datta S, Halloran M, Longini I. Augmented HIV vaccine trial designs for estimating reduction in infectiousness and protective efficacy. Stat Med. 1998; 17:185–200. [PubMed: 9483728]

Datta S, Halloran M, Longini I. Efficiency of estimating vaccine efficacy for susceptibility and infectiousness: Randomization by individual versus household. Biometrics. 1999; 55:792–798. [PubMed: 11315008]

Frangakis C, Rubin D. Principal stratification in causal inference. Biometrics. 2002; 58:21–29. [PubMed: 11890317]

Gilbert P, Bosch R, Hudgens M. Sensitivity analysis for the assessment of causal vaccine effects on viral load in HIV vaccine trials. Biometrics. 2003; 59:531–541. [PubMed: 14601754]

Halloran, M.; Longini, I.; Struchiner, C. Design and Analysis of Vaccine Studies. Springer; New York: 2010.

Halloran M, Préziosi M, Chu H. Estimating vaccine efficacy from secondary attack rates. J Am Stat Assoc. 2003; 98:38–46.

Halloran M, Struchiner C. Study designs for dependent happenings. Epidemiology. 1991; 2:331–338. [PubMed: 1742381]

Halloran M, Struchiner C, Longini I. Study designs for different efficacy and effectiveness aspects of vaccination. Am J Epidemiol. 1997; 146:789–803. [PubMed: 9384199]

Hudgens M, Halloran M. Causal vaccine effects on binary postinfection outcomes. J Am Stat Assoc. 2006; 101:51–64. [PubMed: 19096723]

Hudgens M, Halloran M. Towards causal inference with interference. J Am Stat Assoc. 2008; 103:832–842. [PubMed: 19081744]

Imai K. Sharp bounds on the causal effects in randomized experiments with “truncation-by-death”. Statistics and Probability Letters. 2008; 78:144–149.

Jemiai Y, Rotnitzky A, Shepherd B, Gilbert P. Semiparametric estimation of treatment effects given baseline covariates on an outcome measured after a post-randomization event occurs. J R Stat Soc Series B. 2007; 69:879–901.

Kendrick P, Eldering G. A study in active immunization against pertussis. Am J Hyg, Sect B. 1939; 38:133.

Koopman J, Little R. Assessing HIV vaccine effects. Am J Epidemiol. 1995; 142:1113–20. [PubMed: 7485056]

Longini I, Hudgens M, Halloran M, Sagatelian K. A Markov model for measuring vaccine efficacy for both susceptibility to infection and reduction in infectiousness for prophylactic HIV-1 vaccines. Stat Med. 1999; 18:53–68. [PubMed: 9990692]

Millar E, Watt J, Bronsdon M, Dallas J, Reid R, Santosham M, O'Brien K. Indirect effect of 7-valent pneumococcal conjugate vaccine on pneumococcal colonization among unvaccinated household members. Clin Infect Dis. 2008; 47:989–996. [PubMed: 18781875]

Moulton L, O'Brien K, Kohberger R, Chang I, Reid R, Weatherholtz R, Hackell J, Siber G, Santosham M. Design of a group-randomised *Streptococcus pneumoniae* vaccine trial. Contr Clin Trials. 2001; 22:438–452.

Orenstein W, Bernier R, Hinman A. Assessing vaccine efficacy in the field: Further observations. Epidemiol Rev. 1988; 10:212–241. [PubMed: 3066628]

PHLS Epidemiologic Research Laboratory. Efficacy of pertussis vaccination in England. Br Med J. 1982; 285:357–9. [PubMed: 6284296]

Préziosi M, Halloran M. Effects of pertussis vaccination on transmission: vaccine efficacy for infectiousness. Vaccine. 2003; 21:1853–1861. [PubMed: 12706669]

Rida W. Assessing the effect of HIV vaccination on secondary transmission. Statistics in Medicine. 1996; 15:2393–2404. [PubMed: 8931209]

Rosenbaum, P. Design of Observational Studies. Springer; New York: 2010.

Shepherd B, Gilbert P, Jemiai Y, Rotnitzky A. Sensitivity analyses comparing outcomes only existing in a subset selected post-randomization, conditional on covariates, with application to HIV. Biometrics. 2006; 62:332–342. [PubMed: 16918897]

Sobel M. What do randomized studies of housing mobility demonstrate? Causal inference in the face of interference. J Am Stat Assoc. 2006; 101:1398–1407.

Tchetgen Tchetgen E, VanderWeele T. On causal inference in the presence of interference. Stat Methods Med Res. 2010 DOI: 10.1177/0962280210386779, published online 10 November 2010, in press.

Trollfors B, Taranger J, Lagergard T, Sundh V, Bryla D, Schneerson R, Robbins J. Immunization of children with pertussis toxoid decreases spread of pertussis within the family. Pediatr Infect Dis J. 1998; 17:196–99. [PubMed: 9535245]

VanderWeele T, Tchetgen Tchetgen E. Bounding the infectiousness effects in vaccine trials. Epidemiology. 2011; 22:686–693. [PubMed: 21753730]

Zhang J, Rubin D. Estimation of causal effects via principal stratification when some outcomes are truncated by “death”. J Educ Behav Statist. 2003; 28:353–368.

Table 1

Motivating example of observed data from a randomized study of 11,000 transmission units with primary case of vaccine status *r* and exposed individual with vaccine status *s,* where 0 denotes control and 1 denotes vaccine. In one quarter of the transmission units, both individuals received vaccine, in one quarter, both received control, and in one half of the transmission units, one received vaccine, the other control. In 2500 transmission units, no one was infected.

Primary case with vaccine status *r*	Exposed individual with vaccine status *s*	No. of primary cases	No. of secondary transmissions	SAR_{rs}
0	0	2500	2250	0.90
1	0	500	250	0.45
0	1	4000	2800	0.70
1	1	1500	525	0.35

Table 2

Basic principal stratification with 11 feasible $\mathbf{S}_i = (\mathbf{S}_{i1},\mathbf{S}_{i2})$ vectors and the feasible $\mathbf{Y}_{i2}$ vectors of potential outcomes under the assumptions stated in Section 2. $\mathbf{Y}_{i1}$ vectors not shown due to lack of space.

Basic Principal Stratum		$S_{i1}(z_i)^{\dagger}$	$S_{i2}(z_i)$	$Y_{i2}(z_i)$				
	Prop.	(00) (01) (10) (11)	(00) (01) (10) (11)	(00)	(01)	(10)	(11)	
1	a	0000	0000	*	*	*	*	
2	b	0000	1010	*	*	*	*	
3	b	1100	0000	0	0	*	*	
				1	0	*	*	
				1	1	*	*	
4	c	0100	1010	*	0	*	*	
				*	1	*	*	
5	c	1100	0010	0	0	*	*	
				1	0	*	*	
				1	1	*	*	
6	d	1100	0011	0	0	*	*	
				1	0	*	*	
				1	1	*	*	
7	e	0100	1011	*	0	*	*	
				*	1	*	*	
8	e	1101	0010	0	0	*	0	
				1	0	*	0	
				1	1	*	0	
				0	0	*	1	↑
				1	0	*	1	↑
				1	1	*	1	
9	d	0101	1010	*	0	*	0	
				*	1	*	0	
				*	0	*	1	↑
				*	1	*	1	
10	f	0000	1111	*	*	*	*	

Int J Biostat. Author manuscript; available in PMC 2013 January 06.

Basic Principal Stratum		$S_{i1}(z_i)^{\dagger}$	$S_{i2}(z_i)$	$Y_{i2}(z_i)$				
	Prop.	**(00) (01) (10) (11)**	**(00) (01) (10) (11)**	**(00)**	**(01)**	**(10)**	**(11)**	
11	f	1111	0000	0	0	0	0	
				1	0	0	0	
				0	0	1	0	↑
				1	1	0	0	
				1	0	1	0	
				0	0	1	1	↑
				1	0	1	1	↑
				1	1	1	0	
				1	1	1	1	

The symbol * denotes undefined transmission outcomes in $\mathbf{Y}_{i2}$.

↑ denotes types of transmission units in which vaccination in the primary case enhances transmission.

Prop. is the proportion in each basic principal stratum under the assumptions in the text.

†The four digits under $S_{i1}(\mathbf{z}_i)$ and $S_{i2}(\mathbf{z}_i)$ represent the vector of potential outcomes under the four treatment assignments. For example in principal stratum 5, 1100 indicates S_{i1} (00) = 1, S_{i1} (01) = 1, S_i (10) = 0, S_{i1} (11) = 0, and 0010 indicates S_{i2} (00) = 0, S_{i2} (01) = 1, S_i (10) = 1, S_{i1} (11) = 0.

Table 3

The relation of observed combinations of $(S_{i1}^{obs}, S_{i2}^{obs})$ and assignment vector $\mathbf{Z}_i$ and the basic principal strata in Table 2.

$\mathbf{Z}_i$	$\left(S_{i1}^{obs}, S_{i2}^{obs}\right)$	Y_{i1}^{obs}	Y_{i2}^{obs}	Principal Strata
(00)	(0,0)	*	*	1
(00)	(1,0)	*	0/1	3,5,6,8,11
(00)	(0,1)	0/1	*	2,4,7,9,10
(01)	(0,0)	*	*	1,2
(01)	(1,0)	*	0/1	3,4,5,6,7,8,9,11
(01)	(0,1)	0/1	*	10
(10)	(0,0)	*	*	1,3
(10)	(1,0)	*	0/1	11
(10)	(0,1)	0/1	*	2,4,5,6,7,8,9,10
(11)	(0,0)	*	*	1,2,3,4,5
(11)	(1,0)	*	0/1	8,9,11
(11)	(0,1)	0/1	*	6,7,10

Table 4

Example of observed number of transmission units by vaccine status and by status of primary case in a study of 11,000 transmission units of size two assuming the $Z_{11},\ldots, Z_{N1}, Z_{12},\ldots, Z_{N2}$ are independent with $\Pr[Z_{ij} = 1] = 0.50$ for all i,j. The $j = 1$ and $j = 2$ are ordered such that if there is a primary case transmission unit, it is $j = 1$.

	Number of Transmission Units Z_i		
	Z_i = (00) (S_{i1}^{obs},S_{i2}^{obs})	**Z_i = (01) or Z_i = (10) (S_{i1}^{obs},S_{i2}^{obs})**	**Z_i = (11) (S_{i1}^{obs},S_{i2}^{obs})**
No primary case	**250 (0,0)**	1000 (0,0)	1250 (0,0)
With primary case	2500 (1,0)	4000, unvac index	1500 (1,0)
		500, vac index	
Total	2750	5500	2750

The boldface indicates direct information about the proportion in principal strata under the assumptions.

Table 5

Basic principal stratification based on the potential infection outcomes $\mathbf{S}_{i1} = (S_{i1}(00), S_{i1}(10))$ of individual $j = 1$ with potential postinfection transmission strata based on the infection outcomes in $j = 2$, $\mathbf{Y}_{i2} = (Y_{i2}(00), Y_{i2}(10))$ in the simplest scenario where assumptions (46), (47), and monotonicity hold.

Potential infection strata				**Potential postinfection transmission strata**		
Basic Principal Stratum		**Potential infection outcomes $S_{i1}(z_i)$**		**Potential postinfection transmission outcomes $Y_{i2}(z_i)$**		**Postinfection transmission interpretation**
	Prop.	**(00)**	**(10)**	**(00)**	**(10)**	
Immune	a	0	0	*	*	always undefined
Protected	d	1	0	0	*	
				1	*	
Doomed	f	1	1	0	0	never transmits
				1	0	
				0	1	↑
				1	1	always transmits

* denotes undefined transmission outcomes in $\mathbf{Y}_{i2}$

↑ denotes types of transmission units in which vaccination in the primary case enhances transmission.

Table 6

Basic principal stratification based on the potential infection outcomes $\mathbf{S}_i = (S_{i1},(00), S_{i1}(01), S_{i1}(10), S_{i1}(11))$ of individual $j = 1$ with potential postinfection transmission strata based on the infection outcomes in $j = 2$, $\mathbf{Y}_{i2} = (Y_{i,2}(00), Y_{i,2}(01), Y_{i,2}(10), Y_{i,2}(11)$ if assumption (46) that only individual 1 is exposed outside the transmission unit holds.

Potential infection strata						**Potential postinfection transmission strata**				
Basic Principal Stratum		**Potential infection outcomes $S_{i1}(z_i)$**				**Potential postinfection transmission outcomes**				**Postinfection transmission interpretation**
	Prop.	**(00)**	**(01)**	**(10)**	**(11)**	**(00)**	**(01)**	**(10)**	**(11)**	
Immune	a	0	0	0	0	*	*	*	*	always undefined
Protected	d	1	1	0	0	0	0	*	*	
						1	0	*	*	
						1	1	*	*	
Doomed	f	1	1	1	1	0	0	0	0	never transmits
						0	0	1	0	↑
						0	0	1	1	↑
						1	0	0	0	
						1	1	0	0	
						1	0	1	0	
						1	0	1	1	↑
						1	1	1	0	
						1	1	1	1	always transmits

* denotes undefined transmission outcomes in $\mathbf{Y}_{i2}$.

↑ denotes types of transmission units in which vaccination in the primary case enhances transmission.

Table 7

Basic principal stratification with six feasible vectors of potential outcomes $\mathbf{S}_{i1} = (\mathbf{S}_{i1},\mathbf{S}_{i2})$ with their potential $\mathbf{Y}_{i2} = (\mathbf{Y}_{i1},\mathbf{Y}_{i2})$ transmission outcomes, both can be exposed outside the transmission unit, but only $j = 1$ assigned either vaccine or control.

Basic Principal Stratum		$S_{i1}(\mathbf{Z}_i)$		$S_{i2}(\mathbf{Z}_i)$		$Y_{i1}(\mathbf{Z}_i)$		$Y_{i2}(\mathbf{Z}_i)$		
	Prop.	(00)	(10)	(00)	(10)	(00)	(10)	(00)	(10)	
1	a	0	0	0	0	*	*	*	*	
2	b	0	0	0	1	*	0	*	*	
						*	1	*	*	
3	c	1	0	0	0	*	*	0	*	
						*	*	1	*	
4	d	1	0	0	1	*	0	0	*	
						*	0	1	*	
						*	1	0	*	
						*	1	1	*	
5	e	0	0	1	1	0	0	*	*	
						1	0	*	*	
						1	1	*	*	
6	f	1	1	0	0	*	*	0	0	
						*	*	0	1	↑
						*	*	1	0	
						*	*	1	1	

* denotes undefined transmission outcomes in $\mathbf{Y}_{Ii1}$ and $\mathbf{Y}_{Ii2}$.

↑ denotes types of transmission units in which vaccination in the primary case enhances transmission.

Table 8

Example of observed distribution of transmission units by status of the primary cases $\left(S_{i1}^{obs}, S_{i2}^{obs}\right)$ under assumptions in Appendix B given underlying principal strata of $\mathbf{S}_i$ as defined in Table 2.

Principal Stratum	Prop.	Number of Transmission Units	Z_i			
			Z_i = (00) $(S_{i1}^{obs}, S_{i2}^{obs})$	Z_i = (01) $(S_{i1}^{obs}, S_{i2}^{obs})$	Z_i = (10) $(S_{i1}^{obs}, S_{i2}^{obs})$	Z_i = (11) $(S_{i1}^{obs}, S_{i2}^{obs})$
1	a	1000	**250 (0,0)**	250 (0,0)	250 (0,0)	250 (0,0)
2	b	1000	250 (0,1)	250 (0,0)	250 (0,1)	250 (0,0)
3	b	1000	250 (1,0)	250 (1,0)	250 (0,0)	250 (0,0)
4	c	1000	250 (0,1)	250 (1,0)	250 (0,1)	250 (0,0)
5	c	1000	250 (1,0)	250 (1,0)	250 (0,1)	250 (0,0)
6	d	1000	250 (1,0)	250 (1,0)	250 (0,1)	250 (0,1)
7	e	1000	250 (0,1)	250 (1,0)	250 (0,1)	250 (0,1)
8	e	1000	250 (1,0)	250 (1,0)	250 (0,1)	250 (1,0)
9	d	1000	250 (0,1)	250 (1,0)	250 (0,1)	250 (1,0)
10	f	1000	250 (0,1)	**250 (0,1)**	250 (0,1)	250 (0,1)
11	f	1000	250 (1,0)	250 (1,0)	**250 (1,0)**	250 (1,0)
Total		11,000	2750	2750	2750	2750
Observed data:						
If j = 1, j = 2 distinguished:						
$(S_{i1}^{obs}, S_{i2}^{obs})$						
(0,0)		2500	**250 (0,0)**	500 (0,0)	500 (0,0)	1250 (0,0)
(1,0)		4250	1250 (1,0)	2000 (1,0)	**250 (1,0)**	750 (1,0)
(0,1)		4250	1250 (0,1)	**250 (0,1)**	2000 (0,1)	750 (0,1)
Total		11,000	2750	2750	2750	2750
Observed data:						
if j =1, j = 2 not distinguished:						
			250 (0,0)	1000 (0,0)		1250 (0,0)
			2500 (1,0)	4000, unvac index		1500 (1,0)
				500, vac index		
Total			2750	5500		2750

Prop. is the proportion in each principal stratum under the assumptions in Section 2.