Interference and Sensitivity Analysis

Tyler J. VanderWeele[1], Eric J. Tchetgen Tchetgen[1], M. Elizabeth Halloran[2]
[1] Departments of Epidemiology and Biostatistics, Harvard School of Public Health
[2] Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center and Department of Biostatistics, University of Washington
March 19, 2014

Abstract. Causal inference with interference is a rapidly growing area. The literature has begun to relax the "no-interference" assumption that the treatment received by one individual does not affect the outcomes of other individuals. In this paper we briefly review the literature on causal inference in the presence of interference when treatments have been randomized. We then consider settings in which causal effects in the presence of interference are not identified, either because randomization alone does not suffice for identification, or because treatment is not randomized and there may be unmeasured confounders of the treatment-outcome relationship. We develop sensitivity analysis techniques for these settings. We describe several sensitivity analysis techniques for the infectiousness effect which, in a vaccine trial, captures the effect of the vaccine of one person on protecting a second person from infection even if the first is infected. We also develop two sensitivity analysis techniques for causal effects in the presence of unmeasured confounding which generalize analogous techniques when interference is absent. These two techniques for unmeasured confounding are compared and contrasted.

Key words: Causal inference; infectiousness effect; interference; sensitivity analysis; spillover effect; stable unit treatment value assumption; vaccine trial.

1. Introduction

Cox (1958, p. 19) wrote that there is no interference between different units if the observation on one unit is unaffected by the particular assignment of treatment to the other units. The assumption of no interference is a key component of Rubin's "stable unit treatment value assumption", called SUTVA (Rubin, 1986), that is often required for potential outcomes to be well-defined. However, in many settings, the assumption of no interference obviously does not hold. Consider an individual who, if not vaccinated, would have infected another person, but who, if vaccinated, would not infect that other person. In this case, the infection outcome of the second person depends on the treatment of the first individual, and there is thus interference. Under the assumption of no interference, the effect of a treatment compares two potential outcomes the individual would exhibit under treatment and control. With interference, an individual could have many potential outcomes depending on the treatments assigned to the other individuals (Rubin 1978, 1990).

In some settings, interference is a nuisance while in other settings it creates effects of scientific, public health, or social science interest. An example of the former includes agricultural experiments where treatments in neighboring plots can interfere with one another (Kempton 1997). Fallow rows between treatment plots can sometimes eliminate interference between plots, but more often the interference must be taken into account. In infectious diseases, interference is inherent in the biology of transmission, it cannot be eliminated, and it produces intrinsically interesting effects. Social interaction is a primary source of interference in studies with humans subjects and often cannot be eliminated.

Progress in causal inference with interference has been made recently in different contexts, including those in the social sciences, econometrics, and infectious diseases. Several causal effects can be defined in the presence of interference, and sometimes similar effects have different names in different contexts. Social scientists have long been interested in the effects of neighborhoods on the economic, sociological, and psychological well-being of their inhabitants, resulting in the term neighborhood effects (Sobel 2006). The consequences of interference between individuals in this context are also known as spillover effects. In infectious diseases, these effects were generally called indirect effects of interventions (Halloran and Struchiner 1991).

In Section 1.1 we present informally some examples of studies on causal inference with interference in different contexts. In Section 2, we present formal definitions of direct, indirect, total and overall effects as well as infectiousness effects in the presence of interference. In Section 3 we develop a number of new sensitivity analysis techniques in settings in which causal effects are not identified, either because the effect estimand itself relies on assumptions beyond randomization or because treatment is not randomized and there may be unmeasured confounding. The sensitivity analysis techniques help address these issues of identification in these settings. Section 3 contains the new results of the paper and, as will be seen below, many of these new results in the context of interference build on approaches of Robins et al. (2000) outside the context of interference. Section 4 offers some concluding remarks on directions for future research on interference. A reader who is primarily interested in the technical development can skip Section 1.1 and move on directly to Section 2.

1.1. Motivating examples

1.1.1. Interference and housing mobility

Sobel (2006) considered interference in the Moving to Opportunity (MTO) demonstration sponsored by the U.S. Department of Housing and Urban Development. In this housing mobility experiment in poor neighborhoods in five cities, eligible ghetto residents were randomly assigned to receive one of two forms of relocation assistance or no assistance (control). Sobel argued that the no interference assumption is not plausible for the MTO demonstration because many of the participants likely knew other participants at each of the five sites. Thus, the participants could have influenced each other through social interaction. For example, a family that decided to move to a new neighborhood could give rise to worse outcomes for a family that stayed in the original neighborhood because of the decline in social support for the family that stayed.

Sobel (2006) defined causal estimands and estimators for indirect/spillover effects for the MTO randomized trial of housing vouchers, taking compliance into account. He assumed that interference could occur within the sites, but not across sites, which he called partial interference. He made a key contribution in proposing causal estimands for assessing effects in the presence of interference by averaging causal effects over all possible treatment assignments for a particular allocation strategy compared to a benchmark strategy wherein no units received the treatment assignment. Although his language is different, he essentially defined causal estimands analogous to the direct, indirect, total and overall effects defined in the next section.

He then compared his causal estimands to what is usually estimated in studies of housing mobility not taking interference into account. He showed that what is usually estimated actually gives the difference between (i) the average effect of the voucher on those who received them and (ii) the average effect on those not receiving vouchers of having people leave the neighborhood. Both effects could be negative (detrimental) with the difference positive, thus making it important to take potential interference into account.

1.1.2. Interference in vaccination programs

Motivated by an interest in the effects of vaccination and vaccination programs, Struchiner, Halloran, Robins and Spielman (1990) and Halloran and Struchiner (1991, 1995) conceptually defined direct, indirect, total and overall effects in the presence of interference. The direct effect of a treatment on an individual was defined as the difference between the potential outcome for that individual given treatment compared to the potential outcome for that individual without treatment if the treatment assignment in the others in the population was held fixed. In contrast to direct effects, an indirect effect describes the effect on an individual of the treatment received by others in the group when that individual's treatment was held fixed. In particular, the indirect effect of a treatment on an individual was defined as the difference between the potential outcomes for that individual without treatment when the group (i) receives an intervention program and (ii) receives a benchmark program of no intervention. Total effects describe the combination of direct and indirect effects of a particular treatment assignment on an individual. The total effect of a treatment on an individual is the difference between the potential outcomes for that individual (i) with treatment when the group receives an intervention program and (ii) without treatment when the group re-

Table 1: Illustrative example of a two-stage randomized placebo-controlled cholera vaccine trial based on data from Ali et al. (2005). Group assignment corresponds to 50% or 30% vaccine coverage (from Hudgens and Halloran 2008).

	Group assignment	Vaccine recipients		Placebo recipients	
Group	(% vaccinated)	Total	Cases	Total	Cases
1	50	12541	16	12541	18
2	50	11513	26	11513	54
3	30	10772	17	25134	119
4	30	8883	22	20727	122
5	30	5627	15	13130	92

ceives no intervention. Overall effects describe the average effect of an intervention relative to no intervention.

Halloran and Struchiner (1995) proposed individual-level causal estimands of direct, indirect, total, and overall in the presence of interference by letting the potential outcomes for any individual depend on the vector of treatment assignments to other individuals in the group (Rubin 1978, 1990). However, they did not propose population level causal estimands.

A number of studies have been conducted to estimate indirect, total, or overall effects of vaccination programs outside of the causal inference framework. In the United Kingdom, the indirect effect of a new program of meningococcal C vaccination was estimated by comparing the attack rates in unvaccinated children and adolescents before and after introduction of the program (Ramsay et al. 2003). The United Kingdom introduced routine meningococcal serogroup C vaccination for infants in November 1999. The vaccine was also offered to all children and adolescents aged <18 years in a phased catch-up program. Adolescents were vaccinated first and the program was completed by the end of 2000. About 75% of the children and adolescents were vaccinated. The attack rate in *unvaccinated* infants through adolescents per 100,000 unvaccinated population in July 1998–June 1999 was 4.08 (95% CI 3.7, 4.5) and in July 2001–June 2002 was 1.36 (95% CI 0.86, 1.85). Vaccinating about 75% of the children and adolescents thus seemed to produce an indirect effect, with a relative reduction in the number of confirmed meningococcal C cases in the unvaccinated children and adolescents, of 67% (95% CI: 52, 77).

To obtain group- and population-level causal estimands for direct, indirect, total, and overall causal effects of treatment, Hudgens and Halloran (2008) proposed a two-stage randomization scheme, the first stage at the group level, the second at the individual level within groups based on Sobel's approach of averaging over all possible treatment assignments. As did Sobel (2006), they assumed interference can occur within groups but not across groups. The causal estimands defined by Hudgens and Halloran (2008) are applicable to other situations with interference in fixed groups of individuals where treatment can be assigned to individuals within groups. A brief formal development is given in Section 2.

As an example, Hudgens and Halloran (2008) presented a hypothetical two-stage randomized placebo-controlled trial of cholera vaccines (Table 1). Suppose in the first stage

five geographically separate groups were randomized so two were assigned to vaccinate 50% and three were assigned to vaccinate 30% of individuals, then individuals were randomly assigned to be vaccinated or not. Causal effect estimates (estimated variance) are given in the change in number of cases per 1000 individuals per year. The estimated indirect effect of vaccinating 50% versus 30% in the unvaccinated individuals is 2.81 (3.079). This suggests that vaccinating 50% of the population would result in 2.8 fewer cases per 1000 unvaccinated people per year compared with vaccinating only 30%. Similarly, the estimated total effect is 4.11 (0.672). This suggests that vaccinating 50% of the population would result in 4.1 fewer cases per 1000 vaccinated people per year compared with vaccinating only 30%. The estimated overall effect is 2.37 (1.430). The estimated overall effect is a summary comparison of the two strategies, suggesting that, on average, 50% vaccine coverage results in 2.4 fewer cases of cholera per 1000 individuals per year compared to 30% vaccine coverage. A public health professional could use these estimates in evaluating the cost-benefit of vaccinating more people and preventing more cases versus vaccinating fewer people. The direct effect under 30% coverage is 3.64 (0.178), nearly three times greater than the direct effect under 50% coverage, which is 1.30 (0.856). The difference shows that even the direct effects can depend on the level of coverage due to interference between individuals.

1.1.3. Interference in the context of kindergarten retention

Hong and Raudenbush (2006) considered interference in the context of the effect on reading scores of children of being retained in kindergarten versus being promoted to the first grade. Interference was assumed possible through the dependence of the potential outcomes of reading test scores of one child on whether other children were retained or not. Hong and Raudenbush were principally interested in the effect of a child's being retained and how this varied with being in schools with low retention and versus those with high retention. They used a sample of data from 1080 schools with 471 kindergarten retainees and 10,255 promoted students. In their application, students are clustered in schools. Individual treatment assignment was whether a student is retained. They used a school-level scalar function based on the proportion of the students that were retained to determine whether a school was a 'high-retention' or 'low-retention' school. The study was observational at two levels: schools were not randomized to have high or low retention, and students were not randomized to be retained. However, they framed their analysis within a two-stage randomization procedure similar to that described in Hudgens and Halloran (2008) in which both stages would have been randomized. They also assumed interference was possible within schools but not across schools.

Using a propensity-score based approach, accounting for interference, and assuming that assignment at both the school and the individual level was ignorable given a number of observed individual-level, school-level, and school-aggregated-individual level characteristics, Hong and Raudenbush (2006) obtained estimates of the effect on reading scores of retention in high-retention and low retention schools. Specifically, in low-retention schools, they estimated the effect on reading scores of a student being retained versus being promoted, was -8.18 ($95\%CI: -10.02, -6.34$), and in high retention schools the effect estimate was -8.86 ($95\%CI: -11.56, -6.16$). A standard deviation in reading test scores in this sample is 13.48 points. We will return to this example below to demonstrate sensitivity analysis in

the context of interference.

1.1.4. Interference between two sides of the face

Rosenbaum (2007) took a different approach to causal inference with interference when analyzing randomized experiments than those in previous sections. He pointed out that if Fisher's null hypothesis of no effect for any individual in the population is true, then there is no effect and consequently no interference. Thus, Fisher's permutation test of no effect will have the correct level, even if, under the alternative hypothesis, there would be interference. He presented several examples, including data from a randomized, double-blind experiment in which 15 people received different preparations of botulinum A exotoxin on each side of their face to treat wrinkles to test which was less painful.

Rosenbaum presented exact nonparametric methods for inverting randomization tests to obtain confidence intervals for assessing treatment effect assuming nothing about the structure of the interference between units. He assumed that there were a number of blocks (groups) with a number of individuals within each group, some of which, but not all, were randomized to a treatment, the others to control. He developed a general notation that allowed interference across blocks and did not assume a two-stage randomization. Rosenbaum (2007) differentiated two null hypotheses. The first null hypothesis is that treatment has no primary effect, that is, the response of each unit does not vary under different randomization assignments in the collection of the possible assignment matrices with fixed number randomized in each block. Analogous to the benchmark allocation of Sobel (2006) and the two-stage randomized trials described above where possibly some communities receive only the control intervention, Rosenbaum (2007) invoked uniformity trials in which individuals within treatment groups would be randomly assigned to treatment and control, but everyone in control groups would receive just control. The second null hypothesis is that treatment has no effect, that is, under different randomization assignments in the collection of the possible assignment matrices with fixed number randomized in each block, each individual's response equals his response in a uniformity trial. If there is no effect, then no benefit is gained from receiving the treatment. If there is no primary effect, there is no advantage to being one of the treated individuals, but the benefits could be shared by all of the individuals.

Rosenbaum gave conditions using distribution-free tests in which without performing the uniformity trials, he was able to get confidence statements about the magnitude of the effect and/or primary effect, though not able to distinguish between them. In the botox example, the 15 people are the blocks, the two sides of the face the individuals. All 15 people reported less pain from the treatment containing alcohol. Using his method, the hypotheses of no effect and no primary effect were rejected with a one-sided significance level 0.000031.

Luo et al (2012) extended this approach in the context of a cognitive neuroscience experiment in which the brains of a moderate number of subjects are studied using functional magnetic resonance imaging while challenged with a rapid fire sequence of randomized stimuli. Interference was assumed to occur between units of time in the same individual

1.1.5. Interference and infectiousness effects

In vaccine contexts, a vaccinated person who becomes infected might have a lower probability of transmitting to a susceptible person during a contact than an unvaccinated person

who becomes infected. This is called the effect of the vaccine on infectiousness. In a study in Niakhar, Senegal, for example, Préziosi and Halloran (2003) estimated the relative reduction in infectiousness to household contacts of a vaccinated case of pertussis compared to an unvaccinated case to be 67% (95%CI 29,86). Estimating reduction in infectiousness can be of considerable public health interest, particularly with vaccines that do not protect well against infection.

Developing general methods for causal inference for infectiousness effects poses complicated challenges. Even if the vaccine is randomized, the infectiousness effect is measured only in people who become infected, a post-randomization variable, so the estimate would in general be subject to selection bias. VanderWeele and Tchetgen Tchetgen (2011b), and Halloran and Hudgens (2012a,2012b) proposed causal quantities corresponding to the infectiousness effect in the simple situation of households of size two. The general approach combines causal inference with interference with principal stratification (Frangakis and Rubin, 2002). The latter accounts for the fact that the comparison in the groups who become infected may be subject to selection bias. The causal infectiousness effect is not identifiable without further assumptions. In Section 2.6, we present the bounds that were developed previously. In Section 3.2, we present new results for sensitivity analyses for causal infectiousness effects.

1.1.6. Other approaches

Manski (2012) studied identification of potential outcome distributions when treatment response may have social interactions. He called the no interference assumption the *individualistic treatment response* to differentiate it from other forms of treatment response that depend on social interaction. VanderWeele et al. (2012a) discussed the relation between causal interactions and interference and how under randomization it is possible to test for specific forms of interference. They show that the theory for causal interactions provides a conceptual apparatus for assessing interference as well.

2. Formalization

In this section we present previously developed formalizations of the direct, indirect, total and overall effects as well as the infectiousness effects as background for the development of the new sensitivity analyses under interference in Section 3.

2.1. Notation

Suppose there are $N \geq 1$ groups of individuals, or blocks of units. For $i = 1, \ldots, N$, let n_i denote the number of individuals in group i and let $\mathbf{Z}_i = (Z_{i1}, \ldots, Z_{in_i})$ denote the treatments those n_i individuals receive. Assume Z_{ij} is a dichotomous random variable having values 0 or 1 such that $\mathbf{Z}_i$ can take on 2^{n_i} possible values. Let $\mathbf{Z}_{i(j)}$ denote the $n_i - 1$ subvector of $\mathbf{Z}_i$ with the jth entry deleted. The vector $\mathbf{Z}_i$ is referred to as an intervention or treatment *program*, to distinguish it from the individual treatment Z_{ij}. Let $\mathbf{z}_i$ and z_{ij} denote possible values of $\mathbf{Z}_i$ and Z_{ij}. Define R^j to be the set of vectors of possible treatment programs of length j, for $j = 1, 2, \ldots, n_i$. For example, $R^2 = \{(0,0), (0,1), (1,0), (1,1)\}$.

Denote the potential outcome of individual j in group i under treatment $\mathbf{z}_i$ as $Y_{ij}(\mathbf{z}_i)$. Denote $\mathbf{Y}_i(\mathbf{z}_i)$ as the vector of such outcomes under treatment $\mathbf{z}_i$ for group i. The notation $Y_{ij}(\mathbf{z}_i)$ allows for the possibility that the potential outcome for the individual j may depend

on another individual's treatment assignment in group i, i.e., it allows for interference between individuals within a group. The $Y_{ij}(\mathbf{z}_i)$ potential responses can be assumed fixed, since they do not depend on the realized random assignment of treatments $\mathbf{Z}_i$, whereas the observed responses $Y_{ij}(\mathbf{Z}_i)$ do depend on $\mathbf{Z}_i$ and thus are random variables. We also consider potential outcomes $\mathbf{Y}_i(\mathbf{z}_i)$ that are independent and identically distributed across blocks. Partial interference is assumed to hold, that is, the outcome of one individual can depend on treatment of other individuals in the same block, but not those in different blocks. The form of the interference within groups is assumed unknown and can be of arbitrary form.

2.2. Treatment assignment mechanisms

Following Hudgens and Halloran (2008) consider a two-stage randomization scheme, the first stage at the group level, the second at the individual level within groups. Let ψ and ϕ denote parameterizations that govern the distribution of $\mathbf{Z}_i$ for $i = 1, \ldots, N$. Corresponding to the first stage of randomization, let $\mathbf{S} \equiv (S_1, \ldots, S_N)$ denote the group assignments with $S_i = 1$ if the group is assigned to ψ and 0 if assigned to ϕ. Let ν denote the parameterization that governs the distribution of $\mathbf{S}$ and let $C \equiv \sum_i S_i$ denote the number of groups assigned ψ. Following Sobel (2006), Hudgens and Halloran (2008) focused on a mixed group and mixed individual assignment strategy, whereby a fixed number of groups were allocated to ψ, and within each group, a fixed number of individuals were allocated to treatment versus control. VanderWeele and Tchetgen Tchetgen (2011a) and Tchetgen Tchetgen and VanderWeele (2012) considered what we call a simple randomization scheme whereby treatment is randomly assigned to different individuals within group i according to a Bernoulli probability mass function. The causal estimands defined below have the same form under either randomization scheme, though the different randomization schemes result in subtle differences of interpretation.

2.3. Average potential outcomes

Causal estimands are typically defined in terms of averages of potential outcomes which are identifiable from observable random variables. Following this approach, the potential outcomes for individual j in group i under $z_{ij} = z$ can be written

$$Y_{ij}(\mathbf{z}_{i(j)}, z_{ij} = z), \tag{1}$$

for $z = 0, 1$. Because (1) depends on $\mathbf{z}_{i(j)}$, following Sobel (2006), Hudgens and Halloran (2008) defined the *individual average potential outcome* for individual j in group i under $z_{ij} = z$ by

$$\overline{Y}_{ij}(z; \psi) \equiv \sum_{\omega \in \{0,1\}^{n_i - 1}} Y_{ij}(\mathbf{z}_{i(j)} = \omega, z_{ij} = z) \mathrm{Pr}_{\psi}(\mathbf{Z}_{i(j)} = \omega | Z_{ij} = z).$$

In other words, the individual average potential outcome is the conditional expectation of $Y_{ij}(\mathbf{Z}_i)$ given $Z_{ij} = 1$ under assignment strategy ψ. In contrast, under the simple allocation strategy of VanderWeele and Tchetgen Tchetgen (2011a), the potential outcomes are averaged over the unconditional distribution of $\mathbf{Z}_{i(j)}$. Averaging over individuals, define the *group average potential outcome* under treatment assignment z as $\overline{Y}_i(z; \psi) \equiv \sum_{j=1}^{n_i} \overline{Y}_{ij}(z; \psi)/n_i$.

Finally, averaging over groups, define the *population average potential outcome* under treatment assignment z as $\overline{Y}(z;\psi) \equiv \sum_{i=1}^{N} \overline{Y}_i(z;\psi)/N$. The causal estimands in the next section are defined in terms of the individual, group, and population average potential outcomes. The individual estimands were defined in Halloran and Struchiner (1995), the individual average, group average and population average estimands in Hudgens and Halloran (2008).

2.4. Direct, indirect, total, and overall causal effects

The *individual direct causal effects* of treatment 0 compared to treatment 1 for the individual j in group i were defined by

$$CE_{ij}^{D}(\mathbf{z}_{i(j)}) \equiv Y_{ij}(\mathbf{z}_{i(j)}, Z_{ij} = 1) - Y_{ij}(\mathbf{z}_{i(j)}, Z_{ij} = 0). \tag{2}$$

The *individual average direct causal effect* for the jth individual in the ith group was defined by

$$\overline{CE}_{ij}^{D}(\psi) \equiv \overline{Y}_{ij}(1;\psi) - \overline{Y}_{ij}(0;\psi), \tag{3}$$

i.e., the difference in individual average potential outcomes when $z_{ij} = 1$ and when $z_{ij} = 0$ under ψ. The *group average direct causal effect* as defined by $\overline{CE}_{i}^{D}(\psi) \equiv \overline{Y}_i(1;\psi) - \overline{Y}_i(0;\psi) = \sum_{j=1}^{n_i} \overline{CE}_{ij}^{D}(\psi)/n_i$, and the *population average direct causal effect* by $\overline{CE}^{D}(\psi) \equiv \overline{Y}(1;\psi) - \overline{Y}(0;\psi) = \sum_{i=1}^{N} \overline{CE}_{i}^{D}(\psi)/N$.

The *individual indirect causal effects* of treatment program $\mathbf{z}$ compared with $\mathbf{z}'$ on individual j in group i were defined by

$$CE_{ij}^{I}(\mathbf{z}_{i(j)}, \mathbf{z}'_{i(j)}) \equiv Y_i(\mathbf{z}_{i(j)}, z_{ij} = 0) - Y_i(\mathbf{z}'_{i(j)}, z'_{ij} = 0), \tag{4}$$

where $\mathbf{z}'$ is another n_i dimensional vector of treatment random variables. (Note $\mathbf{z}'$ does not denote the transpose of $\mathbf{z}$). Similar to direct effects, the *individual average indirect causal effect* were defined by $\overline{CE}_{ij}^{I}(\phi,\psi) \equiv \overline{Y}_{ij}(0;\phi) - \overline{Y}_{ij}(0;\psi)$. Clearly if $\psi = \phi$, then $\overline{CE}_{ij}^{I}(\phi,\psi) = 0$; that is, there will be no individual average indirect causal effects. Finally, the *group average indirect causal effect* were defined as $\overline{CE}_{i}^{I}(\phi,\psi) \equiv \overline{Y}_i(0;\phi) - \overline{Y}_i(0;\psi) = \sum_{j=1}^{n_i} \overline{CE}_{ij}^{I}(\phi,\psi)/n_i$. and the *population average indirect causal effect* as $\overline{CE}^{I}(\phi,\psi) \equiv \overline{Y}(0;\phi) - \overline{Y}(0;\psi) = \sum_{i=1}^{N} \overline{CE}_{i}^{I}(\phi,\psi)/N$.

The *individual total causal effects* for individual j in group i were defined as

$$CE_{ij}^{T}(\mathbf{z}_{i(j)}, \mathbf{z}'_{i(j)}) \equiv Y_{ij}(\mathbf{z}_{i(j)}, z_{ij} = 1) - Y_{ij}(\mathbf{z}'_{i(j)}, z'_{ij} = 0). \tag{5}$$

The *individual average total causal effect* was defined by $\overline{CE}_{ij}^{T}(\phi,\psi) \equiv \overline{Y}_{ij}(1;\phi) - \overline{Y}_{ij}(0;\psi)$, the *group average total causal effect* was defined by $\overline{CE}_{i}^{T}(\phi,\psi) \equiv \overline{Y}_i(1;\phi) - \overline{Y}_i(0;\psi) = \sum_{j=1}^{n_i} \overline{CE}_{ij}^{T}(\phi,\psi)/n_i$, and the *population average total causal effect* was defined by $\overline{CE}^{T}(\phi,\psi) \equiv \overline{Y}(1;\phi) - \overline{Y}(0;\psi) = \sum_{i=1}^{N} \overline{CE}_{i}^{T}(\phi,\psi)/N$. It follows by simple addition and subtraction that a total effect is the sum of the direct and indirect effects at the individual, individual average, group average and population average levels. For example: $\overline{CE}^{T}(\phi,\psi) = \overline{Y}(1;\phi) - \overline{Y}(0;\psi) = \overline{Y}(1;\phi) - \overline{Y}(0;\phi) + \overline{Y}(0;\phi) - \overline{Y}(0;\psi) = \overline{CE}_{i}^{D}(\phi) + \overline{CE}^{I}(\phi,\psi)$.

The overall causal effect was defined to be the average effect of an intervention program relative to no intervention. The *individual overall causal effect* of treatment $\mathbf{z}_i$ compared to treatment $\mathbf{z}_i'$ for individual j in group i was defined by $CE_{ij}^O(\mathbf{z}_i, \mathbf{z}_i') \equiv Y_{ij}(\mathbf{z}_i) - Y_{ij}(\mathbf{z}_i')$. Similarly, for the comparison of ϕ to ψ, the *individual average overall causal effect* was defined by $\overline{CE}_{ij}^O(\phi, \psi) \equiv \overline{Y}_{ij}(\phi) - \overline{Y}_{ij}(\psi)$, the *group overall causal effect* by $\overline{CE}_i^O(\phi, \psi) \equiv \overline{Y}_i(\phi) - \overline{Y}_i(\psi)$, and the *population overall causal effect* by $\overline{CE}^O(\phi, \psi) \equiv \overline{Y}(\phi) - \overline{Y}(\psi)$ where $\overline{Y}(\psi) \equiv \sum_{i=1}^N \overline{Y}_i(\psi)/N$ and $\overline{Y}_i(\psi) \equiv \sum_{j=1}^{n_i} \overline{Y}_{ij}(\psi)/n_i$ and $\overline{Y}_{ij}(\psi) \equiv \sum_{\omega \in \{0,1\}^{n_i}} Y_{ij}(\mathbf{z}_i = \omega)\Pr_\psi(\mathbf{Z}_i = \omega)$.VanderWeele and Tchetgen Tchetgen (2011a) showed that the overall effect decomposes into the sum of an indirect effect and a contrast of two direct effects on the individual average, group average and population average levels. For example, $\overline{CE}^O(\phi, \psi) = \overline{CE}^I(\phi, \psi) + \{\overline{CE}^D(\phi)\Pr_\phi(Z_{ij} = 1) - \overline{CE}^D(\psi)\Pr_\psi(Z_{ij} = 1)\}$.

The quantities defined above under interference have two important distinctions from those used in causal inference without interference. First they quantify causal effects only for participants in the randomized study. Second, they depend on the randomization probabilities (through $\Pr_\psi(Z_{i(j)} = \omega | Z_{ij} = z)$). Although the causal estimands here depend on the assignment mechanism (e.g. comparing two different proportions vaccinated), we could alternatively compare allocation strategies of always vaccinate versus never vaccinate to recover traditional causal estimands that do not depend on the assignment mechanism.

The estimands defined above simplify under the assumption of no interference between individuals within a group since the potential outcomes of the jth individual in group i can be written as $Y_{ij}(1)$ and $Y_{ij}(0)$. In turn, the individual direct causal effect is no longer dependent on the treatment assignment vector $\mathbf{z}_{i(j)}$ and simply equals $Y_{ij}(1) - Y_{ij}(0)$. The corresponding group average direct causal effect becomes $\sum_{j=1}^{n_i}\{Y_{ij}(1) - Y_{ij}(0)\}/n_i$, i.e., the usual average causal effect estimand. By (4), the individual indirect causal effect equals zero for all individuals assuming no interference. That is, assuming no interference implies the treatment has no indirect effects. Similarly, by (2) the individual total causal effect equals the individual direct causal effect. Likewise, at the group average level, under the no interference assumption the indirect causal effect is zero and the direct causal effect equals the total causal effect.

2.5. Inference and challenges

Assuming the two-stage randomization and mixed allocation strategy, Hudgens and Halloran proposed unbiased estimators for the various population average effects. They provided variance estimates under the assumption of stratified inference, that is, if it matters only how many people are allocated to treatment, not exactly which ones. Tchetgen Tchetgen and VanderWeele (2012) provided conservative variance estimators (i.e. guaranteed to be no smaller than the true variance in expectation), under more general assumptions and provided finite sample confidence intervals for the various effects without the assumption of stratified interference. Liu and Hudgens (2014) further developed large sample randomization inference for the direct, indirect, total, and overall causal effects in the presence of interference when either the number of groups or the number of individuals within groups grows large, but not necessarily both.

2.6. Interference and infectiousness effects

To develop causal estimands for the infectiousness effects presented in Section 1.1.5, we follow the development of VanderWeele and Tchetgen Tchetgen (2011b) and Halloran and Hudgens (2012a). Consider a setting with N households (groups) indexed by $i = 1, ..., N$. Each household consists of two persons indexed by $j = 1, 2$. We let Z_{ij} denote the vaccine status for individual j in household i, where $Z_{ij} = 1$ if the individual received vaccine and $Z_{ij} = 0$ if the individual did not. For each household, $\mathbf{Z}_i = (Z_{i1}, Z_{i2})$ denotes the vaccine status of the two individuals in the household. We let Y_{ij} denote the infection status of individual j in household i after some suitable follow up in the study. We let $Y_{ij}(z_{i1}, z_{i2})$ denote the potential outcome for individual j in household i if the two individuals in that household i had vaccine status of (z_{i1}, z_{i2}); we treat the potential outcome vector $\mathbf{Y}_i(z_{i1}, z_{i2})$ as a random variable that is independent and identically distributed across households.

We assume partial interference, that is, the exposure status of persons in one household in the study do not affect the outcomes of individuals in other study households. The assumption that clusters constitute isolated pairs would be reasonable in a vaccine trial conducted with a relatively small number of households in a very large city so that it is unlikely that the various households in the study would interact with one another. We will assume that the two individuals in each household are distinguishable (e.g. a husband and wife pair) and we will consider a simple randomized experiment in which only one of the two individuals (e.g. the wife) is predetermined to be randomized to receive a vaccine or control and the second person (e.g. the husband) is predetermined to be always unvaccinated. We let $j = 1$ denote the individual who may or may not be vaccinated (e.g. the wife) and $j = 2$ the individual who is always unvaccinated (e.g. the husband). In other settings in which the individual (husband or wife) who is subject to vaccination is itself randomized (i.e. two-stage randomization) the analysis below could be done separately in those households in which the wife was selected for vaccine randomization versus those in which the husband was selected. Halloran and Hudgens (2012a) generalized that case to the situation where either person can be randomized to vaccine or exposed outside the household.

The crude (or net) estimator for the infectiousness effect on the risk difference scale was defined as:

$$E[Y_{i2}|Z_{i1} = 1, Y_{i1} = 1] - E[Y_{i2}|Z_{i1} = 0, Y_{i1} = 1] \tag{6}$$

where the expectation is taken over all households. This is a comparison of the infection rates for individual 2 in the subgroup in which individual 1 was vaccinated and infected versus in the subgroup in which individual 1 was unvaccinated and infected. Even though the vaccine status for individual 1 is randomized, conditioning on a variable that occurs after treatment, e.g., the infection status of individual 1, in effect breaks randomization. The net estimator for the infectiousness effect could be subject to selection bias. We are computing infection rates for individual 2 for subpopulations that may be quite different with respect to individual 1.

Consider a second contrast proposed by VanderWeele and Tchetgen Tchetgen (2011b) and Halloran and Hudgens (2012a):

$$E[Y_{i2}(1, 0) - Y_{i2}(0, 0)|Y_{i1}(1, 0) = Y_{i1}(0, 0) = 1]. \tag{7}$$

This contrast compares the infection status for individual 2 if individual 1 was vaccinated,

$Y_{i2}(1,0)$, versus unvaccinated, $Y_{i2}(0,0)$, but only among the subset of households for whom individual 1 would have been infected irrespective of whether individual 1 was vaccinated i.e. $Y_{i1}(1,0) = Y_{i1}(0,0) = 1$. Such a subgroup is sometimes called a principal stratum (Frangakis and Rubin, 2002). The contrast in (7) is not subject to selection bias, so it can be considered a formal causal contrast for the infectiousness effect.

Unfortunately, we do not know which households fall into the subpopulation in which individual 1 would have been infected irrespective of whether individual 1 was vaccinated. The contrast (7) is, in general, unidentified, even when treatment is randomized, though the observable data do provide some information about (7). Bounds and sensitivity analysis are further facilitated by other assumptions.

Assumption 1. For all i, $Y_{i1}(1,0) \leq Y_{i1}(0,0)$.

Assumption 1, usually called a monotonicity assumption, states there is no one who would be infected if vaccinated but uninfected if unvaccinated. Under Assumption 1, there are three principal strata, or subgroups of households defined by the joint potential infection outcomes of individual 1 under vaccine and control. They are (i) the doomed principal stratum in which individual 1 is infected whether vaccinated or not, (ii) the protected stratum in which individual 1 is infected if unvaccinated and uninfected if vaccinated, and (iii) the immune stratum, in which individual 1 does not become infected whether vaccinated or not. The causal contrast (7) is defined in the doomed principal stratum.

To simplify notation, let $p_v = E[Y_{i2}(1,0)|Y_{i1}(1,0) = Y_{i1}(0,0) = 1]$, $p_u = E[Y_{i2}(0,0)|Y_{i1}(1,0) = Y_{i1}(0,0) = 1]$, $p_1 = E[Y_{i2}|Z_{i1} = 1, Y_{i1} = 1]$ and $p_0 = E[Y_{i2}|Z_{i1} = 0, Y_{i1} = 1]$. The crude (net) infectiousness effect (6) is then just $p_1 - p_0$, and the causal infectiousness effect (7) is $p_v - p_u$. Under randomization and monotonicity, any household where individual 1 is infected if vaccinated must be in the doomed stratum, so $p_v = p_1$. Thus, one component of the causal infectiousness effect (7) is identified.

However, any household where individual 1 becomes infected if unvaccinated could be in the doomed or protected stratum. Thus p_u is not identified without further assumptions. However, under monotonicity, the ratio ρ of the proportion in the protected stratum to the sum of the proportions in the protected and doomed strata is identified by the observed data. Thus, we know what proportion of the households in which individual 1 received control and was infected is in the doomed stratum, just not which ones, so we do not know what proportion of secondary transmissions occurred in the doomed strata. Under Assumption 1, Halloran and Hudgens (2012a, 2012b) derived upper and lower bounds for causal effects on infectiousness that are constrained by the relation in the data between ρ and p_0.

A further possible assumption is

Assumption 2. $E[Y_{i2}(0,0)|Z_{i1} = 0, Y_{i1} = 1] \leq E[Y_{i2}(0,0)|Z_{i1} = 1, Y_{i1} = 1]$.

Assumption 2 states that the average infection rate for individual 2 if both individuals 1 and 2 were unvaccinated would be lower in the subgroup of households for which individual 1 would be infected and unvaccinated than in the subgroup of households for which individual 1 would be infected and vaccinated. The assumption might be thought plausible insofar as the subgroup for which individual 1 was vaccinated and infected might be less healthy than the subgroup for which individual 1 was unvaccinated and infected; thus, if both people are

unvaccinated, individual 2 is more likely to be infected in the first subgroup than in the second.

Under Assumptions 1 and 2, VanderWeele and Tchetgen Tchetgen (2011b) showed the crude contrast in (6) is conservative for the causal contrast in (7) in that $E[Y_{i2}(1,0) - Y_{i2}(0,0)|Y_{i1}(1,0) = Y_{i1}(0,0) = 1] \leq E[Y_{i2}|Z_{i1} = 1, Y_{i1} = 1] - E[Y_{i2}|Z_{i1} = 0, Y_{i1} = 1]$ i.e. $p_v - p_u \leq p_1 - p_0$. Analogous results in fact also hold for the risk ratio, odds ratio, and vaccine efficacy scales (VanderWeele and Tchetgen Tchetgen, 2011b).

The approach may be employed outside of the vaccine context. For example, in an observational study in which the treatment is a smoking cessation program in which one of two persons in a household participated. The participation of the first person might affect the smoking behavior of the second. This might occur either (i) because smoking cessation for the first person encourages the second to stop smoking or because (ii) even if the first person does not stop smoking, the second person might nevertheless be exposed to some of the smoking cessation program materials. One could evaluate this second type of effect (the analogue of the infectiousness effect) by applying the approach described above.

3. Interference and Sensitivity Analysis

3.1. *Overview*

In this section we develop sensitivity analysis techniques that can help assess the presence of causal effects in two settings where these effects in the presence of interference are not identified. These causal effects may not be identified either because the treatments are not randomized or because, even if the treatments are randomized, the spillover effects of interest involve conditioning on a post-treatment variable thereby breaking randomization. Building on the previous sections, we first consider the setting of a randomized trial where the spillover effect of interest is not identified by randomization alone because of conditioning on a post-randomization variable as in the infectiousness effect described in Section 2.6. We present sensitivity analysis methods for assessing this infectiousness effect. We then consider the setting of observational data such as in Hong and Raudenbush (2006) in which causal effects and spillover effects may not be identified due to one or more unmeasured confounding variables. We present two sensitivity analysis techniques for causal effects in the presence of interference that extend analogous results for causal effects under no-interference (Robins et al., 2000; VanderWeele and Arah, 2011) to the setting of causal effects and spillover effects in the presence of interference.

3.2. *Sensitivity Analysis for the Infectiousness Effect*

In Section 2.6, we described two previously developed approaches to bounds on the infectiousness effects. Here we develop methods for sensitivity analysis for the infectiousness effect. We follow the development first in VanderWeele and Tchetgen Tchetgen (2011b) and Halloran and Hudgens (2012a) and then in Hudgens and Halloran (2006); further technical development is given in the Appendix. See also VanderWeele and Tchetgen Tchetgen (2011b) and Hudgens and Halloran (2006) for concrete applications. A simple sensitivity analysis approach also follows from the development of VanderWeele and Tchetgen Tchetgen (2011b). We use the same notation as in Section 2.6. As noted in Section 2.6, under monotonicity Assumption 1, we have that $p_v = p_1$, and thus to obtain the causal infectiousness effect we

need to express p_u in terms of the observed data and sensitivity analysis parameters. We will describe three different parameterizations.

First, let $\theta = E[Y_{i2}(0,0)|Z_{i1} = 1, Y_{i1} = 1] - E[Y_{i2}(0,0)|Z_{i1} = 0, Y_{i1} = 1]$ denote the sensitivity parameter which contrasts the average counterfactual infection rates for individual 2 if both individuals 1 and 2 were unvaccinated in the subgroup of households for which individual 1 is vaccinated and infected versus the subgroup of households for which individual 1 is unvaccinated and infected. It follows from the development in VanderWeele and Tchetgen Tchetgen that, under monotonicity, $p_u = p_0 + \theta$ and thus:

$$p_v - p_u = p_1 - p_0 - \theta$$

In other words to obtain the infectiousness effect under monotonicity, we can calculate the crude infectiousness effect in (6), specify the sensitivity parameter θ, and subtract the sensitivity parameter θ from the crude estimate to obtain the infectiousness effect. We can vary θ over a range of plausible values in a sensitivity analysis to produce a range of plausible values for the infectiousness effect. The sensitivity analysis parameter is subject to certain empirical constraints as described below. However, because of the simple relationship above, a corrected confidence interval under sensitivity parameter θ can be obtained simply by subtracting θ from both limits of the confidence interval for the crude estimate in (6).

We can also use a similar approach but with a different parameterization of the sensitivity analysis parameters. Following Hudgens and Halloran (2006), we can vary $\gamma = E[Y_{i2}(0,0)|Y_{i1}(1,0) = 0, Y_{i1}(0,0) = 1]$, the probability of secondary transmission in the protected stratum when individual 1 receives control with bounds set by constraints of the data (Halloran and Hudgens 2012a, 2012b). The quantity γ is not identifiable from the observed data without further assumptions, but once a value of γ is assumed, then the probability of secondary transmission in the doomed stratum is fixed, and thus p_u is identified. Varying γ, the infectiousness effect (7) on the risk difference scale can be obtained as $p_1 - p_u$ where p_u is given by

$$p_u = \frac{p_0 - \gamma(1-\rho)}{\rho} \ ,$$

and where γ can vary between

$$\max\left\{0, \frac{p_0 - \rho}{1-\rho}\right\} \leq \gamma \leq \min\left\{1, \frac{p_0}{1-\rho}\right\},$$

with the left side giving rise to the upper bound for the causal infectiousness effect and the right side giving rise to the lower bound on the risk difference scale. Note that a drawback of this approach is that the parameter γ is constrained by the observed data and similar restrictions are imposed on θ above by virtue of the identity $\theta = p_0 - \frac{p_0 - \gamma(1-\rho)}{\rho}$.

As a third parameterization, we could follow an approach to sensitivity analysis developed in Scharfstein et al. (1999) and Robins et al. (2000). This approach performs sensitivity analysis with a bias parameter β on the ratio scale. Gilbert et al. (2003) and Hudgens and Halloran (2006) adapted this approach for sensitivity analysis for causal effects on post-infection outcomes, the former in the continuous post-infection outcome scenario, the latter for binary post-infection outcomes. Hudgens and Halloran (2006) and Halloran and Hudgens

(2012a) suggested that this approach to sensitivity analysis could be used for infectiousness effects taking as the intermediate infection outcome the infection status of individual 1 and as the potential post-infection outcome the infection status of individual 2. Within the context of the infectiousness effect, again under monotonicity assumption 1, the bias parameter β can be expressed as:

$$\exp(\beta) = \frac{P(Y_{i2}(0,0)=1|Z_{i1}=1,Y_{i1}=1)/P(Y_{i2}(0,0)=0|Z_{i1}=1,Y_{i1}=1)}{P\{Y_{i2}(0,0)=1|Y_{i1}(0,0)=1,Y_{i1}(1,0)=0\}/P\{Y_{i2}(0,0)=0|Y_{i1}(0,0)=1,Y_{i1}(1,0)=0\}}.$$

The bias parameter β is the log of odds ratio comparing the risk of infection if individual 1 is not vaccinated among (i) the doomed stratum and (ii) the protected. Note this is not simply a different scale than the bias parameter θ above (odds ratio versus risk difference) but also a comparison of different subpopulations. Once this bias parameter is specified then it can be shown that $p_u = E[Y_{i2}(0,0)|Y_{i1}(1,0) = Y_{i1}(0,0) = 1]$ is the positive root of $(-b \pm \sqrt{b^2 - 4zc})/2z$, where

$$\begin{aligned} z &= \exp(\beta)p_0 \\ b &= \left(1 - \frac{P(Y_{i1}=1|Z_{i1}=1)}{P(Y_{i1}=1|Z_{i1}=1)}\right)\{\exp(\beta)-1\} - \exp(\beta)p_0 + p_0 - \exp(\beta) \\ c &= \frac{P(Y_{i1}=1|Z_{i1}=1)}{P(Y_{i1}=1|Z_{i1}=1)}\{\exp(\beta)-1\}. \end{aligned}$$

For further discussion of inference for this approach to sensitivity analysis see Hudgens and Halloran (2006) and Jemiai et al. (2007).

In each of the three parameterizations above, once we have obtained $p_u = E[Y_{i2}(0,0)|Y_{i1}(1,0) = Y_{i1}(0,0) = 1]$ we can obtain the infectiousness effect on the difference, risk ratio, odds ratio, or infectiousness effect scales by $p_v - p_u$, p_v/p_u, $p_v(1-p_u)/\{p_u(1-p_v)\}$ and $1 - p_v/p_u$, respectively.

We have focused here on the setting of a randomized trial, but the approach is potentially applicable to observational studies as well if, conditional on some set of covariates C, the treatment was jointly independent of the counterfactual outcomes (i.e. effectively randomized within strata of C). The sensitivity analysis parameters would have to be conditional on C.

3.3. Sensitivity Analysis for Spillover Effects Under Unmeasured Confounding: Approach 1

We now consider a setting in which causal effects and spillover effects under interference are not identified due to unmeasured confounding. Adjustment is often made for covariates to attempt to control for such confounding. However, in an observational study we can never be sure that the control is adequate. One or more unmeasured confounders may bias effect estimates. Confounding control becomes even more complex in settings with interference since when one individual's outcome is under consideration, control will often need to be made for the covariates of other individuals in the same cluster (Tchetgen Tchetgen and VanderWeele, 2012; Ogburn and VanderWeele, 2012, Perez-Heydrich et al 2013). Unmeasured confounding can thus operate either through the unmeasured covariates for the focal individual or for other individuals in the same cluster. In this sub-section, we apply and extend the sensitivity analysis approach of VanderWeele and Arah (2011) to allow for set-

tings with interference and spillover effects. In the next sub-section we consider an extension of the sensitivity analysis approach of Robins et al. (2000) to allow for interference and spillover effects.

We consider a general observational setting such as that employed by Hong and Raudenbush (2006) wherein individuals are clustered in groups such that individuals within groups may influence one another but there is no interference between groups. We make the stratified interference assumption above and further assume, following Hong and Raudenbush, that the potential outcome of person j, $Y_{ij}(\mathbf{z}_i)$, depends on the treatment received by the individuals in cluster i other than person j, $\mathbf{z}_{i(j)}$, only through some known many-to-one scalar function $g(\mathbf{z}_{i(j)})$ so that $Y_{ij}(\mathbf{z}_i)$ can be written as $Y_{ij}\left(z_{ij}, g(\mathbf{z}_{i(j)})\right)$. For example, $g(\mathbf{z}_{i(j)})$ might be the mean of $\mathbf{z}_{i(j)}$. Let $G_{ij} = g(\mathbf{Z}_{i(j)})$. For all i, j, Z_{ij} is determined by simple randomization. We then have that

$$E[Y(z,g)\,|Z = z, G = g] = E[Y(z,g)].$$

Hong and Raudenbush (2006) considered a variation on this assumption in the context of observational data. Specifically, for some covariate vector L_{ij}, they assumed that

$$E[Y(z,g)\,|Z = z, G = g, L = l] = E[Y(z,g)\,|L = l] \tag{8}$$

and from this it follows that

$$E[Y(z,g)\,|L = l] = E[Y|Z = z, G = g, L = l]$$

where the right hand side can be estimated with observed data. Hong and Raudenbush (2006) also allowed L_{ij} to contain cluster level covariates along with cluster aggregates of individual level covariates. Note, however, that (8) requires that $Y_{ij}(z,g)$ be mean independent of both Z_{ij} and $g(\mathbf{Z}_{i(j)})$ conditional on L_{ij}. If, for each individual, Z_{ij} is randomized conditional on L_{ij}, although this will imply that $Y_{ij}(z,g)$ is mean independent of Z_{ij} conditional on L_{ij}, it does not necessarily guarantee that $Y_{ij}(z,g)$ is mean independent of $g(\mathbf{Z}_{i(j)})$ conditional on L_{ij}. Let $\mathbf{L}_{i(j)}$ denote the vector of all covariates L_{ij} for all individuals in cluster i other than individual j. We might, instead of (8), consider

$$E[Y(z,g)\,|Z = z, G = g, L = l, h(\mathbf{L}) = h] = E[Y(z,g)\,|L = l, h(\mathbf{L}) = h] \tag{9}$$

where $h(\mathbf{L}_{i(j)})$ is a known function of $\mathbf{L}_{i(j)}$. However once again, with (9), even if, for each individual, Z_{ij} were randomized conditional on $L_{ij}, h(\mathbf{L}_{i(j)})$, this would not guarantee that $Y_{ij}(z,g)$ is mean independent of $g(\mathbf{Z}_{i(j)})$ conditional on $L_{ij}, h(\mathbf{L}_{i(j)})$ unless $h(\mathbf{L}_{i(j)}) = \mathbf{L}_{i(j)}$. See Ogburn and VanderWeele (2012) for discussion of causal structures for which assumptions (8) or (9) will hold. Under assumption (9), we have

$$E[Y(z,g)\,|L = l, h(\mathbf{L}) = h] = E[Y|Z = z, G = g, L = l, h(\mathbf{L}) = h]$$

where again the right hand side can be estimated with observed data. From this one could

obtain conditional direct, indirect and total effects, namely,

$$E[Y(z,g)|l,h] - E[Y(z',g)|l,h]$$
$$E[Y(z,g)|l,h] - E[Y(z,g')|l,h]$$
$$E[Y(z,g)|l,h] - E[Y(z',g')|l,h].$$

These contrasts are important insofar as they allow one to assess the relative importance for an individual's outcome of changing an individual's own treatment versus the treatment of other individuals. In other words, the effects allow one to assess the relative importance of spillover. Marginal effects, involving counterfactuals of the form $E[Y(z,g)]$ could be obtained by averaging over the distributions of L_{ij} and $h(\mathbf{L}_{i(j)})$.

Suppose now that we have unmeasured confounding by one or more unmeasured confounders U_{ij} and let $\mathbf{U}_{i(j)}$ denote the vector of U_{ij} for all individuals in cluster i other than individual j. Suppose that the analogue of assumption (9) holds conditional on observed $L_{ij}, h(\mathbf{L}_{i(j)})$ and unobserved $U_{ij}, v(\mathbf{U}_{i(j)})$ for some scalar function v so that

$$E[Y(z,g)\,|Z=z, G=g, L=l, h(\mathbf{L})=h, U, v(\mathbf{U})] = E[Y(z,g)\,|L=l, h(\mathbf{L})=h, U, v(\mathbf{U})], \tag{10}$$

but that (9) does not hold when we do not condition on $U_{ij}, v(\mathbf{U}_{i(j)})$. Without data on U_{ij} causal effects are not identified. Let $H = h(\mathbf{L}_{i(j)})$ and $V = v(\mathbf{U}_{i(j)})$.

Following the sensitivity analysis approach of VanderWeele and Arah (2011) for causal effects under no-interference, we express the difference between the causal effect

$$E[Y(z,g)|l,h] - E[Y(z',g')|l,h] = \sum_{u,v}\{E[Y|z,g,l,h,u,v] - E[Y|z',g',l,h,u,v]\}P(u,v|l,h)$$

and the biased estimand

$$E[Y|z,g,l,h] - E[Y|z',g',l,h]$$

in terms of sensitivity analysis parameters. Let $B = \{E[Y|z,g,l,h] - E[Y|z',g',l,h]\} - \{E[Y(z,g)|l,h] - E[Y(z',g')|l,h]\}$ denote this difference. Technical development is given in the appendix.

Let u^* and v^* denote arbitrary reference values for U and V, respectively. Under assumption (10) we have that:

$$\begin{aligned} B \;=\; & \sum_{u,v}\{E(Y|z,g,l,h,u,v) - E(Y|z,g,l,h,u^*,v^*)\}\{P(u,v|z,g,l,h) - P(u,v|l,h)\} \\ & - \sum_{u,v}\{E(Y|z',g',l,h,u,v) - E(Y|z',g',l,h,u^*,v^*)\}\{P(u,v|z',g',l,h) - P(u,v|l,h)\}. \end{aligned} \tag{11}$$

To obtain the bias factor B one could thus specify the effect of the unmeasured confounders U and V on the outcome, $E(Y|z,g,l,h,u,v) - E(Y|z,g,l,h,u^*,v^*)$, for $(Z,G) = (z,g)$ and $(Z,G) = (z',g')$, and also how the distribution of U and V differs when $(Z,G) = (z,g)$ versus $(Z,G) = (z',g')$, i.e., $P(u,v|z,g,l,h)$ and $P(u,v|z',g',l,h)$. One can use these sensitivity analysis parameters to calculate the bias factor in (11) and then subtract the bias factor B from the estimate of the causal effect using the observed data $E[Y|z,g,l,h] - E[Y|z',g',l,h]$

to obtain a corrected effect estimate for $E[Y(z,g)|l,h] - E[Y(z',g')|l,h]$.

Note that the expression for the bias factor in (11) makes no assumption beyond assumption (10) that control for observed (L,H) and unobserved (U,V) would suffice to control for confounding of the effect of (Z,G) on Y; it allows for multiple unmeasured confounders. However, the use of the bias formula in (11) requires specifying a large number of parameters: $E(Y|z,g,l,h,u,v) - E(Y|z,g,l,h,u^*,v^*)$ for every value of u,v and the distributions $P(u,v|z,g,l,h)$ and $P(u,v|z',g',l,h)$.

Under some simplifying assumptions expression (11) reduces to a much easier to use formula. In particular, suppose that there is a single unmeasured confounder U and that $V = v(\mathbf{U}_{i(j)})$ is scalar. Suppose also that the effects of U and $V = v(\mathbf{U}_{i(j)})$ are additive in the sense that $E(Y|z,g,l,h,u,v) - E(Y|z,g,l,h,u^*,v^*) = \lambda(u-u^*) + \tau(v-v^*)$ for $(Z,G) = (z,g)$ and $(Z,G) = (z',g')$. In the appendix it is shown that under these assumptions:

$$B = \lambda\{E[U|z,g,l,h] - E[U|z',g',l,h]\} + \tau\{E[V|z,g,l,h] - E[V|z',g',l,h]\}. \tag{12}$$

To use this simplified bias formula one only needs to specify the effect, λ, for a one unit increase in the unmeasured confounder U_{ij}, the effect τ of a one unit increase in the scalar functional of the unmeasured confounders of the other members of the group, $v(\mathbf{U}_{i(j)})$, and how the means of U_{ij} and $v(\mathbf{U}_{i(j)})$ differ when $(Z,G) = (z,g)$ versus when $(Z,G) = (z',g')$. Once these sensitivity analysis parameters are specified the bias factor B can be calculated using formula (12) above and then B could be subtracted from the estimate of the causal effect using the observed data $E[Y|z,g,l,h] - E[Y|z',g',l,h]$ to obtain a corrected effect estimate for $E[Y(z,g)|l,h] - E[Y(z',g')|l,h]$. Under this simplified approach because the bias factor involves only the sensitivity analysis parameters and not the observed data, a corrected confidence interval could be obtained by subtracting B from both limits of a confidence interval for $E[Y|z,g,l,h] - E[Y|z',g',l,h]$. A sensitivity analysis consists of reporting the causal contrasts under a range of plausible values of τ and λ. The values $\tau = \lambda = 0$ correspond to the assumption of no unmeasured confounders. Selecting a plausible range of parameters could be done using subject matter expertise, by external data from other studies, or by including a very wide range of parameters that are thought to include those that would constitute very extreme values. Alternatively, one could examine the most important measured confounder and assess the magnitude of the corresponding parameters for the most important measured confounder; one could then consider whether an additional unmeasured confounder with parameters set equal to that of the most important measured confounder would substantially alter results. This would allow an investigator to assess whether an unmeasured confounder would have to be stronger than the most important measured confounder to substantially alter the results.

Consider again the substantive example of Hong and Raudenbush (2006) described in Section 1.1.3. As noted above, Hong and Raudenbush (2006) examined the effect of kindergarten retention on reading test scores allowing for interference by allowing the retention of other students at the school to affect a child's reading test scores. They assumed that treatment assignment at both the school and the individual level was ignorable given a number of observed individual-level, school-level, and school-aggregated-individual level characteristics. Using a propensity-score based approach they estimated, in the notation above, the contrasts $E[Y(z=1,g=1)] - E[Y(z=0,g=1)]$ and $E[Y(z=1,g=0)] - E[Y(z=0,g=0)]$ where

$g = 1$ denotes high-retention school and $g = 0$ a low retention school. They found that, in low-retention schools, their estimates indicated that the effect on reading scores of a student being retained versus being promoted, was -8.18 $(95\%CI : -10.02, -6.34)$, and in high retention schools the effect estimate was -8.86 $(95\%CI : -11.56, -6.16)$. A standard deviation in reading test scores in this sample is 13.48 points.

Hong and Raudenbush also went through a sensitivity analysis argument for their results. They noted that the strongest predictor of current test scores were lagged test scores, but that it was unlikely that there was any unmeasured covariate that would predict their outcomes so strongly. They considered instead whether unmeasured individual and school covariates that had effects on readings scores that were equal to those of the measured covariates with second strongest association with reading scores would suffice to explain away the effect estimates. Using an argument based on a formula similar to (12), they reported that unmeasured individual and school confounders that had an effect as large as the second most important measured individual and school level covariates would shift the estimate in high retention schools to -4.25 $(95\%CI : -6.95, -1.54)$ and thus not suffice to bring the confidence interval to include 0. However, in low-retention schools, unmeasured individual and school confounders that had an effect as large as the second most important measured covariates would shift the confidence interval in low retention schools to -0.60 $(95\%CI : -2.44, 1.24)$ and thus would suffice to bring their confidence interval for the effect in low retention schools below 0. The effects in high retention schools seem more robust to the possibility of unmeasured confounding. In their analyses, Hong and Raudenbush (2006) used a similar expression to (12), but in their paper they did not provide a derivation of this formula and did not articulate the assumptions needed for the use of the formula. We have provided the derivations and assumptions required here. Moreover, we have also provided a more general expression, that in (11), that is applicable under much weaker assumptions.

We have considered here a sensitivity analysis approach for causal effects and spillover effects in the presence of interference. In other contexts, questions concerning whether the effect of a treatment on an outcome is mediated by some intermediate may be of interest. In settings in which mediation is of interest and interference occurs at the level of the mediator so that the mediator for one unit may affect the outcomes for other units (cf. VanderWeele, 2010a; VanderWeele et al., 2013), a similar sensitivity analysis approach for unmeasured confounding of the mediator and the outcome could be developed by applying and extending the results of VanderWeele (2010b) for direct and indirect effects from the no-interference setting to a setting with interference by following an analogous approach to that presented above.

3.4. Sensitivity Analysis for Spillover Effects Under Unmeasured Confounding: Approach 2

In this sub-section we consider an alternative sensitivity analysis approach to assess the influence of unobserved confounding for direct and spillover effects in general settings similar to those considered by Hong and Raudenbush (2006) described above. The following developments follow closely from analogous sensitivity analysis techniques recently proposed in the context of mediation analysis (Tchetgen Tchetgen, 2011; Tchetgen Tchetgen and Shpitser, 2012). In order to formalize the approach, suppose that we wish to make inferences

about the following causal effects:

$$\gamma^d(z,g,l,h) = E[Y(z,g) - Y(z_0,g)\,|z,g,l,h]$$
$$\gamma^s(g,l,h) = E[Y(z_0,g) - Y(z_0,g_0)\,|g,l,h]$$

where, unless stated otherwise, throughout, G and H are left unrestricted, i.e. $G_{i(j)} = g(\mathbf{Z}_{i(j)}) = \mathbf{Z}_{i(j)}, H_{i(j)} = h(\mathbf{L}_{i(j)}) = \mathbf{L}_{i(j)}$. These effects are versions, allowing for interference, of the so-called effect of treatment on the treated, which have been studied extensively in the absence of interference, by econometricians, epidemiologists and social scientists. The first contrast $\gamma^d(z,g,l,h)$ captures the direct effect of Z_{ij} on Y_{ij} conditional on the person's observed exposure Z_{ij}, and the cluster's observed data $(\mathbf{Z}_{i(j)}, L_{ij}, H_{i(j)})$. In contrast, $\gamma^s(g,l,h)$ is the spillover causal effect of $\mathbf{Z}_{i(j)}$ on $Y_{ij}(z_{i,j} = z_0)$ within levels of $(\mathbf{Z}_{i(j)}, L_{ij}, H_{i(j)})$. Note that $\gamma^d(z_0,g,l,h) = \gamma^s(g_0,l,h) = 0$, so that these effects are relative to the reference average potential outcome under (z_0, g_0).

Consider the pair of no unobserved confounding assumptions:

$$\begin{aligned} E[Y(z_0,g)\,|z,g,l,h] &= E[Y(z_0,g)\,|z_0,g,l,h] \\ E[Y(z_0,g_0)\,|g,l,h] &= E[Y(z_0,g_0)\,|g_0,l,h] \end{aligned}$$

It is straightforward to show that under these assumptions, γ^d and γ^s are identified by:

$$\gamma^{d,\dagger}(z,g,l,h) = E[Y|z,g,l,h] - E[Y|z_0,g,l,h]$$
$$\gamma^{s,\dagger}(g,l,h) = E[Y|z_0,g,l,h] - E[Y|z_0,g_0,l,h]$$

This shows that the above no unobserved confounding assumption suffices to identify direct and spillover effects (on the treated) using a standard regression analysis approach to model $E[Y|z,g,l,h]$. Next, suppose that the no unobserved confounding assumption does not hold, we define the following selection bias functions:

$$\delta^d(z,g,l,h) = E[Y(z_0,g)\,|z,g,l,h] - E[Y(z_0,g)\,|z_0,g,l,h] \tag{13}$$
$$\delta^s(g,l,h) = E[Y(z_0,g_0)\,|g,l,h] - E[Y(z_0,g_0)\,|g_0,l,h] \tag{14}$$

where, $\delta^d(z_0,g,l,h) = \delta^s(g_0,l,h) = 0$. These selection bias functions come about naturally upon contrasting, on the additive scale, each of the observational conditional association $\gamma^{d,\dagger}(z,g,l,h)$ and $\gamma^{s,\dagger}(g,l,h)$, with their corresponding causal analog, $\gamma^d(z,g,l,h)$ and $\gamma^s(g,l,h)$, respectively. To illustrate in the simple context of binary Z, one can verify that

the confounding bias quantified on the additive scale is equal to

$$\begin{aligned}
& \gamma^{d,\dagger}(z=1,g,l,h) - \gamma^{d}(z=1,g,l,h) \\
= \; & E[Y|z=1,g,l,h] - E[Y|z_0,g,l,h] \\
& -E\left[Y(z=1,g)\,|z=1,g,l,h\right] + E\left[Y(z_0,g)\,|z=1,g,l,h\right] \\
= \; & E[Y(z=1,g)\,|z=1,g,l,h] - E[Y(z_0,g)\,|z_0,g,l,h] \\
& -E\left[Y(z=1,g)\,|z=1,g,l,h\right] + E\left[Y(z_0,g)\,|z=1,g,l,h\right] \\
= \; & E\left[Y(z_0,g)\,|z=1,g,l,h\right] - E[Y(z_0,g)\,|z_0,g,l,h] \\
= \; & \delta^{d}(z=1,g,l,h)
\end{aligned}$$

which makes clear the central role of the selection bias function δ^d. A generalization of the above derivation gives similar expressions for $\delta^s(g,l,h)$, and also extends beyond binary Z. Furthermore, this derivation also makes clear that the presence of confounding implies that at least one of the following must hold:

$$\begin{aligned}
\text{either } \delta^{d}(z,g,l,h) &\neq 0 \text{ for some } (z,g,l,h) \\
\text{or } \delta^{s}(g,l,h) &\neq 0 \text{ for some } (g,l,h).
\end{aligned}$$

The first condition implies γ^d is not identified, while the second case implies γ^s is not identified. Thus, we may proceed as in Robins et al (2000), and recover causal inferences by assuming the selection bias functions, $\delta^d(z,g,l,h)$ and $\delta^s(g,l,h)$, that encode the magnitude and direction of unmeasured confounding, are known. Suppose that higher values of Y are beneficial to one's health. If $\delta^d(1,g,l,h) > 0$, then on average, an individual j in cluster i with $\left\{G_{i(j)} = g, L_{ij} = l, H_{i(j)} = h\right\}$ and exposure value $Z_{ij} = 1$ has higher potential outcomes $Y_{ij}(z_0 = 0, g)$ than an individual in the same cluster and the same stratum $\left\{G_{i(j)} = g, L_{ij} = l, H_{i(j)} = h\right\}$ but unexposed $Z_{ij} = 0$; i.e., healthier individuals are more likely to receive the exposure conditional on the exposures of other people in the cluster and the observed confounders for the cluster. On the other hand, $\delta^d(1,g,l,h) < 0$ suggests confounding by indication for exposure; i.e. unhealthier individuals are more likely to be exposed. Likewise, if $\delta^s(g,l,h) > 0$ for all $g \neq g_0$ indicates that on average, an individual in a cluster with confounders $\left\{ L_{ij} = l, H_{i(j)} = h\right\}$ and $g(\mathbf{Z}_{i(j)}) = g^* \neq g_0$ has higher potential outcomes $Y_{ij}(z_0 = 0, g_0)$ than a comparable individual in a comparable cluster with baseline exposure value $g(\mathbf{Z}_{i(j)}) = g_0$. In the special case where $g\left(\mathbf{Z}_{i(j)}\right) = \mathbf{Z}_{i(j)} = \mathbf{0}$, one has that clusters with no exposed individual tend on average, to be less healthy than clusters with one or more individuals exposed.

The approach to inference in the presence of confounding involves the following reparameterization of the conditional mean function $E[Y|z,g,l,h]$ in terms of the causal contrasts γ^d and γ^s, and the selection bias functions δ^d and δ^s. To state the reparameterization, suppose for the moment that $f(z,g|l,h) = f\left(\mathbf{Z}_i = (z,g)\,|\mathbf{L}_i = (l,h)\right)$, is known, then one can verify

that for each unit in cluster i,

$$
\begin{aligned}
&E[Y|z,g,l,h] \\
&= \gamma^d(z,g,l,h) + \delta^d(z,g,l,h) - \sum_{z=0}^{1} \delta^d(z,g,l,h)\, f(z|g,l,h) \\
&\quad + \gamma^s(g,l,h) + \delta^s(g,l,h) - \sum_{\mathbf{z}^* \in \{0,1\}^{n_i-1}} \delta^s(g(\mathbf{z}^*),l,h)\, f(\mathbf{z}^*|l,h) + q(l,h)
\end{aligned}
$$

where

$$q(l,h) \equiv E[Y(z_0,g_0)|l,h].$$

For fixed δ^d and δ^s given by equations (13) and (14), the causal contrasts γ^d and γ^s are nonparametrically identified and can be estimated by fitting the above regression model.

In practice, due to the high dimensionality of $\mathbf{L}_i$ often encountered in applications, parametric models must be used to reliably estimate $\gamma^d(z,g,l,h), \gamma^s(g,l,h), f(z,g|l,h)$ and $q(l,h)$. The above reparameterization is particularly advantageous in that it ensures variation independence of parameters of working models for these various quantities. A description of parametric maximum likelihood and generalized estimating equations estimation is given in the appendix.

3.5. A comparison of sensitivity analysis techniques

It is instructive to compare the sensitivity analysis techniques given in Sections 3.3 and 3.4. We begin by noting that the causal estimand targeted by the two methods differ. The first approach aims to make inferences about

$$E[Y(z,g) - Y(z',g)|l,h] \text{ and } E[Y(z',g) - Y(z',g')|l,h],$$

while the second approach targets the causal contrasts

$$E[Y(z,g) - Y(z',g)|z,g,l,h] \text{ and } E[Y(z',g) - Y(z',g')|g,l,h].$$

In the absence of confounding, the contrasts targeted by the two approaches coincide, but when unmeasured confounding is present, the first approach gives direct and spillover effects for the subset of individuals with $\{l,h\}$, while the second approach delivers inferences about the direct effect for individuals with $\{z,g,l,h\}$ and the spillover effect for individuals with $\{g,l,h\}$, i.e. interference effects of treatment on the treated. This distinction has implications for the corresponding sensitivity analysis techniques. Although both approaches require specifying unidentified parameters, in the first technique the parameters correspond to particular causal effects (of U and V), whereas in the second technique the parameters do not correspond to causal effects. More specifically, in the first approach, a parametrization of the bias expression (11) involves quantifying the causal interaction

$$
\begin{aligned}
&\{E(Y|z,g,l,h,u,v) - E(Y|z,g,l,h,u^*,v^*)\} - \{E(Y|z',g',l,h,u,v) - E(Y|z',g',l,h,u^*,v^*)\} \\
=\ &\{E(Y|z,g,l,h,u,v) - E(Y|z',g',l,h,u,v)\} - \{E(Y|z,g,l,h,u^*,v^*) - E(Y|z',g',l,h,u^*,v^*)\}
\end{aligned}
$$

of the effect of (Z,G) within levels of (L,H,U,V). The first sensitivity analysis approach

requires not only making a judgement about the nature of U (i.e. binary, polytomous, multivariate, etc.), and the magnitude and the direction of unmeasured confounding, but also about the magnitude and direction of effect heterogeneity on the additive scale. In contrast, the second sensitivity analysis technique directly quantifies the magnitude and direction of unmeasured confounding without making any reference to a specific U, and therefore it does not involve making any judgement about unidentified causal effects. While making a judgement about the nature of U (i.e. binary, polytomous, multivariate, etc.) in the first approach could be difficult in practice, the second approach, to ensure that posited models are compatible, requires that the user posit parametric models for $f(z|g,l,h)$ and $f(\mathbf{G}=g|l,h)$, an additional modeling requirement not needed by the first approach. Both of these densities are, however, nonparametrically identified from the observed data and, therefore, standard goodness-of-fit tools may be adopted to ensure a reasonable fit to the data. The first approach can be of particular use if subject matter expertise can help determine the nature of the unmeasured confounders and/or if prior analyses with other data for which these confounders are available can be used to help inform the value of the sensitivity analysis parameters.

The difference between the two techniques also has interesting but subtle implications if these techniques are used to construct tests of the sharp null of no treatment effects. If an investigator were interested in testing the sharp null of no treatment effects using the first sensitivity analysis approach, care would need to be taken to ensure that the specification of the sensitivity analysis parameters was compatible with the sharp null e.g. by ensuring the interaction function in the above display is 0 as, for example, in the simplified expression in (12). In contrast, with the second sensitivity analysis approach, any specification of the sensitivity analysis parameters will be compatible with the sharp null, and so this issue of compatibility is not similarly a concern.

4. Discussion

In this paper we have reviewed definitions and approaches to causal inference in the presence of interference. We have developed various sensitivity analysis approaches when the causal effects and spillover effects of interest are unidentified either because randomization does not suffice for identification (Hudgens and Halloran, 2006; VanderWeele and Tchetgen Tchetgen, 2011b; Halloran and Hudgens, 2012ab) or because of unmeasured confounding in an observational study.

We have extended existing sensitivity analysis approaches (Robins et al., 2000; VanderWeele and Arah, 2011) from the setting of no-interference to allow for interference. Many settings in the social sciences in which causal effects and spillover effects are of interest are observational settings with interference and the results presented here will likely be useful in those settings. Further work could be done on sensitivity analysis for unmeasured confounding in other settings in which there are not multiple independent clusters (e.g. Aronow and Samii, 2013) or in settings involving the assessment of causal effects in social networks (Christakis and Fowler, 2007; VanderWeele, 2011).

More generally, numerous further challenges remain in the development of methods for causal inference under interference. Inference may become additionally challenging for treatment with multiple levels as the number of combinations will increase dramatically. Inter-

ference patterns might also depend on the covariates of individuals in a cluster in complex ways. Furthermore, the approach for defining causal estimands for infectiousness where the covariates of individuals in the group are taken into account becomes unwieldly when the clusters are larger than two, posing an additional challenge for future research. When dealing with cluster-randomized studies in which the clusters are large, the number of clusters may be small, making inference difficult. In contrast, in the household studies, the number of clusters may be large, but the number in the households small.

One of the limitations of the approaches for causal inference with interference presented in this paper is the assumption that there were fixed groups or blocks of individuals. One of the challenges for future research will be the issue of interference across groups. More generally, recent research on causal inference with interference has been relaxing the assumption of fixed groups. Aronow and Samii (2013) presented randomization-based methods for estimating average causal effects under arbitrary interference of known form. Van der Laan (2012) incorporates network information for each individual under consideration that describes the set of other individuals that each individual is potentially connected to. Inference is then driven but the number of individuals rather the number of communities, which in this case is one. Liu and Hudgens (2013) propose new generalized inverse probability weighted estimators of causal effects in the presence of any form of interference between individuals. Defining and comparing causal estimands of effects across more general forms of interference will be challenging. Much more exciting research is left to be done on causal inference under general forms of interference.

Acknowledgements. The authors thank the editors and reviewers for helpful comments. TJV was supported by NIH grant R01 ES017876. EJTT was supported by NIH grants R01ES020337, R01AI104459, R21ES019712 and U54GM088558. MEH was supported by NIH grants R37 AI032042 and R01 AI085073. The content is solely the responsibility of the authors and does not necessarily reflect the official views of the National Institute of Allergy and Infectious Diseases or the National Institutes of Health.

Appendix

Derivations for the Infectiousness Effect

VanderWeele and Tchetgen Tchetgen (2011b) showed that under monotonicity assumption 1, $p_v = E[Y_{i2}(1,0)|Y_{i1}(1,0) = Y_{i1}(0,0) = 1] = E[Y_{i2}|Z_{i1} = 1, Y_{i1} = 1] = p_1$ and $p_u = E[Y_{i2}(0,0)|Y_{i1}(1,0) = Y_{i1}(0,0) = 1] = E[Y_{i2}|Z_{i1} = 0, Y_{i1} = 1] + \{E[Y_{i2}(0,0)|Z_{i1} = 1, Y_{i1} = 1] - E[Y_{i2}(0,0)|Z_{i1} = 0, Y_{i1} = 1]\} = p_1 + \theta$. From this it follows that $p_v - p_u = (p_1 - p_0) - \theta$ and $p_v/p_u = p_1/(p_0 + \theta)$, $p_v(1 - p_u)/\{p_u(1 - p_v)\} = p_1(1 - p_0 - \theta)/\{(p_0 + \theta)(1 - p_1)\}$, and $1 - p_v/p_u = 1 - p_1/(p_0 + \theta)$.

Hudgens and Halloran (2006) showed that under monotonicity assumption 1, $p_u = \gamma B/\{1 + \gamma(B - 1)\}$ and $p_0 = \gamma V + p_u(1 - V)$ where $B = \exp(\beta)$, $\gamma = P\{Y_{i2}(0,0)|Y_{i1}(1,0) = 0, Y_{i1}(0,0) = 1\}$, and $V = \left(1 - \frac{P(Y_{i1}=1|Z_{i1}=1)}{P(Y_{i1}=1|Z_{i1}=1)}\right)$. Solving $p_u = \gamma B/\{1 + \gamma(B - 1)\}$ and $p_0 = \gamma V + p_u(1 - V)$ to eliminate γ gives $p_u = \frac{p_0 - p_u(1-V)}{V} B/\{1 + \frac{p_0 - p_u(1-V)}{V}(B - 1)\}$ or $0 = p_0 B - p_u(-V - Bp_0 + p_0 - B + VB) + (B - 1)(1 - V)p_u^2$, a quadratic equation in p_u. The

roots of this equation are $\frac{-b\pm\sqrt{b^2-4zc}}{2z}$ where $z=\exp(\beta)p_0$, $b=\left(1-\frac{P(Y_{i1}=1|Z_{i1}=1)}{P(Y_{i1}=1|Z_{i1}=1)}\right)\{\exp(\beta)-1\}-\exp(\beta)p_0+p_0-\exp(\beta)$, and $c=\frac{P(Y_{i1}=1|Z_{i1}=1)}{P(Y_{i1}=1|Z_{i1}=1)}\{\exp(\beta)-1\}$.

Derivations for Sensitivity Analysis of Spillover Effect Under Unmeasured Confounding: Approach 1

If in the notation of VanderWeele and Arah (2011), we let A, X and U be (Z,G), (L,H) and (U,V), respectively, and we let a_1 and a_0 denote (z,g) and (z',g'), respectively, then by VanderWeele and Arah (2011) we have that

$$\begin{aligned}
B &= E[Y|z,g,l,h]-E[Y|z',g',l,h]-E[Y(z,g)|l,h]-E[Y(z',g')|l,h] \\
&= E[Y|z,g,l,h]-E[Y|z',g',l,h]-\sum_{u,v}\{E[Y|z,g,l,h,u,v]-E[Y|z',g',l,h,u,v]\}P(u,v|l,h) \\
&= \sum_{u,v}\{E(Y|z,g,l,h,u,v)-E(Y|z,g,l,h,u^*,v^*)\}\{P(u,v|z,g,l,h)-P(u,v|l,h)\} \\
&\quad -\sum_{u,v}\{E(Y|z',g',l,h,u,v)-E(Y|z',g',l,h,u^*,v^*)\}\{P(u,v|z',g',l,h)-P(u,v|l,h)\}.
\end{aligned}$$

If $E(Y|z,g,l,h,u,v)-E(Y|z,g,l,h,u^*,v^*)=\lambda(u-u^*)+\tau(v-v^*)$ then we have

$$\begin{aligned}
B &= \sum_{u,v}\{E(Y|z,g,l,h,u,v)-E(Y|z,g,l,h,u^*,v^*)\}\{P(u,v|z,g,l,h)-P(u,v|l,h)\} \\
&\quad -\sum_{u,v}\{E(Y|z',g',l,h,u,v)-E(Y|z',g',l,h,u^*,v^*)\}\{P(u,v|z',g',l,h)-P(u,v|l,h)\} \\
&= \sum_{u,v}\{\lambda(u-u^*)+\tau(v-v^*)\}\{P(u,v|z,g,l,h)-P(u,v|l,h)\} \\
&\quad -\sum_{u,v}\{\lambda(u-u^*)+\tau(v-v^*)\}\{P(u,v|z',g',l,h)-P(u,v|l,h)\} \\
&= \sum_{u,v}(\lambda u+\tau v)P(u,v|z,g,l,h)-\sum_{u,v}(\lambda u+\tau v)P(u,v|z',g',l,h) \\
&= \lambda\{E[U|z,g,l,h]-E[U|z',g',l,h]\}+\tau\{E[V|z,g,l,h]-E[V|z',g',l,h]\}.
\end{aligned}$$

Maximum Likelihood and GEE Estimation of Direct and Spillover Effects Under Unmeasured Confounding: Approach 2.

We briefly describe maximum likelihood and generalized estimating equations inference for direct and spillover effects in the presence of unobserved confounding under Approach 2. As stated in the text, in practice, due to the high dimensionality of $\mathbf{L}_i$ often encountered in applications, parametric models must be used to reliably estimate $\gamma^d(z,g,l,h)$, $\gamma^s(g,l,h)$, $f(\mathbf{Z}_i|\mathbf{L}_i)$ and $q(l,h)$. The proposed reparameterization of $E[Y|z,g,l,h]$ is particularly advantageous in that it ensures variation independence of parameters of the working models for these various quantities making possible a straightforward application of maximum likelihood estimation. Specifically, consider the parametric models $\gamma^d\left(z,g,l,h;\psi^d\right)$, $\gamma^s(g,l,h;\psi^s)$, $f(\mathbf{Z}_i|\mathbf{L}_i;\alpha)$ and

$q(l,h;\eta)$; then, provided parameters are not shared across working models, a particular choice of one of these models is guaranteed by our parametrization not to place any restriction on the other models. Maximum likelihood estimation of unknown parameters requires that one posit an additional working model for the conditional density $f(\varepsilon_i|\mathbf{L}_i)$ which we denote $f(\varepsilon_i|\mathbf{L}_i;\omega)$, for ε_i the vector of possibly correlated residuals $Y_{ij}-\ E[Y|z,g,l,h]$, $j=1,...,n_i$. In principle, our choice of parametrization could be used in conjunction with standard techniques for modeling clustered outcomes, such as for instance by incorporating a random intercept to introduce correlation within a cluster in the regression model of Y_{ij}. Maximum likelihood estimation then proceeds by maximization of

$$\log\prod_{i=1}^{N} f\left(\varepsilon_i\left(\psi^d,\psi^s,\eta,\alpha\right)|\mathbf{L}_i,z;\omega\right) f\left(\mathbf{Z}_i|\mathbf{L}_i;\alpha\right)$$

with respect to $\left(\psi^d,\psi^s,\eta,\alpha,\omega\right)$. A simple alternative to maximum likelihood estimation entails finding $\widehat{\alpha}$ that maximizes the partial log-likelihood

$$\log\prod_{i=1}^{N} f\left(\mathbf{Z}_i|\mathbf{L}_i;\alpha\right)$$

and then finding the parameter value $\left(\widehat{\psi}^d,\widehat{\psi}^s,\widehat{\eta}\right)$ that solves the following generalized estimating equation with independence working correlation structure:

$$\sum_{ij}\frac{\partial\varepsilon_{ij}\left(\psi^d,\psi^s,\eta,\widehat{\alpha}\right)}{\partial\left(\psi^d,\psi^s,\eta\right)|_{\widehat{\psi}^d,\widehat{\psi}^s,\widehat{\eta}}}\varepsilon_{ij}\left(\widehat{\psi}^d,\widehat{\psi}^s,\widehat{\eta},\widehat{\alpha}\right)=0.$$

Note that the dependence of ε_{ij} on $\left(\delta^d,\delta^s\right)$ has been suppressed in the notation used above; such dependence is made explicit in a sensitivity analysis which is obtained by repeating either maximum likelihood estimation or the estimating equations approach given above as $\left(\delta^d,\delta^s\right)$ is varied within a finite set of user-specified functions $\Gamma=\{\ \delta^d_{\lambda^d},\delta^s_{\lambda^s}:\lambda=\left(\lambda^d,\lambda^s\right)\}$ indexed by a finite dimensional parameter λ with $\left(\delta^d_0,\delta^s_0\right)\in\Gamma$ corresponding to the ignorability assumption, i.e. $\delta^d_0=\delta^s_0\equiv 0$.

In applying this approach it is helpful to briefly describe possible functional forms for the selection bias functions δ^d,δ^s. In practice, it may be convenient to specify simple parametric models for each of these functions as illustrated in the following display. To illustrate, one may set the function $g(\mathbf{Z}_{i(j)})=\sum_{j'\neq j}Z_{ij'}$ to equal the number of exposed individuals in cluster i excluding person j:

$$\begin{aligned}\delta^{d,1}_{\lambda^d}(z,g,l,h) &= \lambda^d z, \qquad \delta^{s,1}_{\lambda^s}(g,l,h)=\lambda^s g,\\ \delta^{d,2}_{\lambda^d}(z,g,l,h) &= z\ \left(\lambda^d_1\ +\lambda^d_2 g\right), \qquad \delta^{s,2}_{\lambda^s}(g,l,h)=g\ \left(\lambda^s_1\ +\lambda^s_2 l\right)\end{aligned}$$

where for $\delta^{d,1}_{\lambda^d}$ and $\delta^{d,2}_{\lambda^d}$, the scalar parameters λ^d and $\left(\lambda^d_1,\lambda^d_2\right)$ encode the magnitude and direction of unmeasured confounding for the effect of person j's exposure Z_{ij} on his out-

come Y_{ij}, and for $\delta^{d,1}_{\lambda^s}$ and $\delta^{d,2}_{\lambda^s}$, the scalar parameters λ^s and $(\lambda^s_1, \lambda^s_2)$ encode the magnitude and direction of unmeasured confounding for the causal effect of the total number exposed $\sum_{j' \neq j} Z_{ij'}$ excluding person j, within the cluster i on person j's outcome Y_{ij}.

The functions $\delta^{d,2}_{\lambda^d}$ and $\delta^{s,2}_{\lambda^s}$ model interactions between Z_{ij} and $\sum_{j' \neq j} Z_{ij'}$, thus allowing for heterogeneity in the selection bias function. Since the functional form of $\left(\delta^d_{\lambda^d}, \delta^s_{\lambda^s}\right)$ is not identified from the observed data, we generally recommend reporting results for a variety of functional forms.

References

Ali, M., Emch, M., von Seidlein, L., Yunus, M., Sack, D.A., Rao, M., Holmgren, J. and Clemens, J.D. (2005). Herd immunity conferred by killed oral cholera vaccine in Bangladesh: a reanalysis. *Lancet* 366:44-49.

Aronow, P.M., and Samii, C. (2013) Estimating average causal effects under general interference. arXiv:1305.6156v1 at arxiv.org/pdf.

Christakis, N.A. and Fowler, J.H. (2007). The spread of obesity in a large social network over 32 years, *New England Journal of Medicine*, 357:370-379.

Cox, D.R. (1958). *The Planning of Experiments*. New York: Wiley.

Frangakis, C.E. and Rubin, D.B. (2002). Principal stratification in causal inference. *Biometrics*, 58:21-29.

Gilbert, P.B., Bosch, R., and Hudgens, M.G. (2003). Sensitivity analysis for the assessment of causal vaccine effects on viral load in HIV vaccine trials. *Biometrics*, 59:531-541.

Halloran, M.E. and Hudgens, M.G. (2012a). Causal inference for vaccine effects on infectiousness. *International Journal of Biostatistics*, 8:(2) Article 6, DOI: 10.2202/1557-4679.1354.

Halloran, M.E. and Hudgens, M.G. (2012b). Comparing bounds for vaccine effects on infectiousness. *Epidemiology*, 23:931-932.

Halloran, M.E. and Struchiner, C.J. (1991). Study designs for dependent happenings. *Epidemiology*, 2:331-338

Halloran, M.E. and Struchiner, C.J. (1995). Causal inference for infectious diseases. *Epidemiology*; 6:142-51.

Hong, G. and Raudenbush, S.W. (2006) Evaluating kindergarten retention policy: A case study of causal inference for multilevel observational data. *Journal of the American Statistical Association*, 101:901-910.

Hudgens, M.G. and Halloran, M.E. (2006) Causal vaccine effects on binary post-infection outcomes. *Journal of the American Statistical Association*, 101:51-64.

Hudgens, M.G. and Halloran, M.E. (2008) Towards causal inference with interference. *Journal of the American Statistical Association*, 103:832-842, 2008.

Jemiai, Y., Rotnitzsky, A., Shepherd, B.E. and Gilbert, P.B. (2007). Semiparametric estimation of treatment effects given base-line covariates on an outcome measured after a post-randomization event occurs. *Journal of the Royal Statistical Society, Series B*, 69:879-902.

Kempton, R.A. (1997) Interference between plots. In R. A. Kempton and P. N. Fox, editors, *Statistical Methods for Plant Variety Evaluation*, pages 101-116, London: Chapman Hall.

Liu, L., and Hudgens, M.G. (2014). Large sample randomization inference of causal effects in the presence of interference. *Journal of the American Statistical Association*, in press

Liu, L., and Hudgens, M.G. (2013). On inverse probability weighted estimators in the presence of interference. Technical Report.

Luo, X., Small, D.S. Li, C.R., and Rosenbaum, P.R. (2012). Inference with interference between units in an fMRI experiment of motor inhibition. *J Am Statist Assoc*, 107:530-541.

Manski, C.F. (2013). Identification of treatment response with social interactions. *Econometrics Journal*,DOI: 10.1111/j.1368-423X.2012.00368.x, 16:S1-S23.

Ogburn, E.L. and VanderWeele, T.J., Causal diagrams for interference and contagion. Submitted.

Perez-Heydrich, C., Hudgens, M.G., Halloran, M.E., Clemens, J.D., Ali, M. and Emch, M.E. (2013) Assessing effects of cholera vaccination in the presence of interference, submitted.

Préziosi, M.-P. and Halloran, M.E. (2003) Effects of pertussis vaccination on transmission: vaccine efficacy for infectiousness, *Vaccine*, **21**:1853–1861.

Ramsay, M.E., Andrews, N.J., Trotter, C.L., Kaczmarski, E.B., Miller, E. (2003) Herd immunity from meningococcal serogroup C conjugate vaccination in England: Database analysis, *Br Med J*, **326**, 365–6.

Robins, J. M., Scharfstein, D., and Rotnitzky, A. (2000). Sensitivity analysis for selection bias and unmeasured confounding in missing data and causal inference models. In: *Statistical Models for Epidemiology, the Environment, and Clinical Trials*. Halloran, E. and Berry, D. (eds), 1-94. New York: Springer-Verlag.

Rosenbaum, P.R. (2007). Interference between units in randomized experiments. *Journal of the American Statistical Association*, 102:191-200.

Rubin, D.B. (1978). Bayesian inference for causal effects: The role of randomization. *Ann Stat*, 7:34-58.

Rubin, D.B. (1986). Which if's have causal answers? *Journal of the American Statistical Association*, 81:961-962.

Rubin, D.B. (1990). Comment: Neyman (1923) and causal inference in experiments and observational studies. *Stat Sci*, 5:472-480.

Scharfstein, D.O., Rotnitzky, A., and Robins, J.M. (1999). Adjusting for non-ignorable drop-out using semiparametric non-response models. *Journal of the American Statistical Association*, 94:1096-1120.

Sobel, M.E. (2006) What do randomized studies of housing mobility demonstrate?: Causal inference in the face of interference. *Journal of the American Statistical Association*, 101:1398-1407.

Struchiner, C.J., Halloran, M.E., Robins, J.M., and Spielman, A. (1990). The behavior of common measures of association used to assess a vaccination program under complex disease transmission patterns - a computer simulation study of malaria vaccines. *Int J Epidemiol*, 19:187-196.

Tchetgen Tchetgen, E.J. and VanderWeele, T.J. (2012). On causal inference in the presence of interference. *Statistical Methods in Medical Research - Special Issue on Causal Inference*, 21:55-75.

Tchetgen Tchetgen, E.J. (2011). Mediation analysis with a survival outcome. *International Journal of Biostatistics*;7(1):Article 33.

Tchetgen Tchetgen, E.J. and Shpitser, I. (2012). Semiparametric theory for causal mediation analysis: efficiency bounds, multiple robustness and sensitivity analysis. *Annals of Statistics*, 40:1816-1845.

Van der Laan, M.J. (2012). Causal inference for networks. UC Berkeley Working Paper Series No. 300.

VanderWeele, T.J. (2010a). Bias formulas for sensitivity analysis for direct and indirect effects. *Epidemiology*, 21:540-551.

VanderWeele, T.J. (2010b). Direct and indirect effects for neighborhood-based clustered and longitudinal data. *Sociological Methods and Research*, 38:515-544.

VanderWeele, T.J. (2011). Sensitivity analysis for contagion effects in social networks. *Sociological Methods and Research*, 40:240-255.

VanderWeele, T.J. and Arah, O.A. (2011). Bias formulas for sensitivity analysis of unmeasured confounding for general outcomes, treatments and confounders. *Epidemiology*, 22:42-52.

VanderWeele, T.J. and Tchetgen Tchetgen, E.J. (2011a). Effect partitioning under interference for two-stage randomized vaccine trials. *Statistics and Probability Letters*, 81:861-869.

VanderWeele, T.J. and Tchetgen Tchetgen, E.J. (2011b). Bounding the infectiousness effect in vaccine trials. *Epidemiology*, 22:686-693.

VanderWeele, T.J., Vandenbroucke, J.P., Tchetgen Tchetgen, E.J., and Robins, J.M. (2012a). A mapping between interactions and interference: implications for vaccine trials. *Epidemiology*, 23:285-292.

VanderWeele, T.J., Hong, G., Jones, S. and Brown, J. (2013). Mediation and spillover effects in group-randomized trials: a case study of the 4R's educational intervention. *Journal of the American Statistical Association*, in press.