HHS Public Access

Author manuscript

Epidemiology. Author manuscript; available in PMC 2015 September 23.

Published in final edited form as:

Epidemiology. 2012 March ; 23(2): 285–292. doi:10.1097/EDE.0b013e318245c4ac.

A mapping between interactions and interference: implications for vaccine trials

Tyler J. VanderWeele,
Departments of Epidemiology and Biostatistics, Harvard School of Public Health

Jan P. Vandenbroucke,
Department of Clinical Epidemiology, Leiden University Medical Center

Eric J. Tchetgen Tchetgen, and
Departments of Epidemiology and Biostatistics, Harvard School of Public Health

James M. Robins
Departments of Epidemiology and Biostatistics, Harvard School of Public Health

Abstract

In this paper we discuss relationships between causal interactions within the counterfactual framework and interference in which the exposure of one person may affect the outcomes of another. We show that the empirical tests for causal interactions can in fact all be adapted to empirical tests for particular forms of interference. In the context of interference, by recoding the response as some function of the outcomes of the various persons within a cluster, a wide range of different forms of interference can potentially be detected. The correspondence between causal interactions and forms of interference extends to encompass *n*-way causal interactions, interference between *n* persons within a cluster and to multi-valued exposures. The theory for causal interactions provides a complete conceptual apparatus for assessing interference as well. The results are illustrated using data from an hypothetical vaccine trial to reason about specific forms of interference and spillover effects that may be present in this vaccine setting. We discuss the implications of this correspondence for our conceptualizations of interaction and for application to vaccine trials.

When considering the joint effects of two exposures, there may be cases in which an outcome occurs if both exposures are present but not if only one or the other is present. We might refer to such settings as instances of "causal interaction".[1–4] This notion is closely related to that of synergism within Rothman's sufficient cause framework.[2,5] Recent work has derived empirical tests[1,2,6] for such interactions. In general, tests for such causal interactions do not correspond to tests for interactions in standard statistical models[2,4,7] unless additional assumptions are made.

There has also been recent interest and methodological development concerning the notion of "interference" or "spillover effects" in which the exposure of one person may affect the

Corresponding Author: Tyler J. VanderWeele, Harvard School of Public Health, Departments of Epidemiology and Biostatistics, 677 Huntington Avenue, Boston MA 02115, Phone: 617-432-7855, Fax: 617-432-1884, tvanderw@hsph.harvard.edu.

outcomes of another person. Such would be the case, say, with "herd immunity" in vaccine trials. The problem of empirical analysis of causal effects in the presence of interference was, until recently, generally considered intractable; however, a number of recent papers[8–14] have shown that progress can in fact be made in certain instances concerning the analysis of such spillover effects.

We show that these two apparently distinct areas of inquiry - causal interaction (concerning the joint effect of two exposures) and interference (spillover effects concerning the exposure of one person affecting the outcomes of another) - are in fact intimately related. In particular we show that an entire battery of empirical tests for casual interaction[1–4,15,16] have a direct mapping onto testing for specific forms of interference.

The remainder of this paper is organized as follows. In the next two sections we review notation, results and concepts for causal interaction and interference. We then describe how the empirical tests for causal interaction can be applied more or less directly to test for specific forms of interference. We consider the case of two persons each possibly receiving a binary treatment or exposure and then in the Appendix generalize to cases of an arbitrary number of exposures or persons and to multi-valued exposures. We illustrate the results for interference using data from an hypothetical vaccine trial. We close with some further discussion on the interrelationships between interaction and interference.

Review of Causal Interactions

Let A_1 and A_2 denote two binary exposures of interest and let D denote a binary outcome. Suppose for now that a person's outcome depends on the values of the two exposures that the person receives, but not on the exposures received by other people. Such an assumption is generally referred to in the statistics literature as one of "no interference". We note that this use of "interference" is different from that in the infectious disease literature where the term connotes interference of the circulation of one infectious agent with the circulation of another. Let $D_{a_1a_2}$ denote the counterfactual outcome (or potential outcome) for D for a person if (possibly contrary to fact) A_1 had been set to a_1 and A_2 had been set to a_2. VanderWeele and Robins[2] said that a sufficient-cause interaction was present between A_1 and A_2 if there was some person such that $D_{11} = 1$ but $D_{10} = D_{01} = 0$, and they showed that if there were such a person then synergism within Rothman's sufficient cause framework[5] would be present. Note that the condition allows D_{00} to be 0 or 1. For such a person the outcome occurs if both exposures are present, but not if just one or the other is present.

We will assume for simplicity that the effects of the exposures A_1 and A_2 on D are unconfounded (in the sense that $D_{a_1a_2}$ is independent of (A_1, A_2)) and that there is no measurement error or selection bias. If the effects of the exposures A_1 and A_2 on D are unconfounded conditional on some set of covariates C, then we will assume that analysis is done within strata of C. Let $p_{a_1a_2} = P(D = 1|A_1 = a_1, A_2 = a_2)$ be the observed probability of the outcome among those who actually received $A_1 = a_1, A_2 = a_2$. The standard test for interaction on the additive scale is

$$p_{11} - p_{10} - p_{01} + p_{00} > 0. \quad (1)$$

VanderWeele and Robins[1,2] showed that if the effects of A_1 and A_2 on D were unconfounded, then a slightly different condition, namely,

$$p_{11} - p_{10} - p_{01} > 0 \quad (2)$$

would imply the presence of a sufficient-cause interaction. The test for a sufficient-cause interaction in condition (2) is a more stringent condition than the standard additive interaction in condition (1), in that we are no longer adding p_{00} in the probability contrast. In fact the magnitude of the contrast $p_{11} - p_{10} - p_{01}$ gives a lower bound on the prevalence of persons with a sufficient cause interaction.

VanderWeele and Robins[1,2] also noted that the test for a standard additive interaction in condition (1) would suffice to conclude the presence of a sufficient-cause interaction under an assumption called "monotonicity." The effects of A_1 and A_2 are said to be positive monotonic if $D_{a_1 a_2}$ is non-decreasing in a_1 and a_2 for all persons i.e. if an increase in the exposure A_1 or A_2 would increase or leave unchanged the outcome, not just on average, but for all people in the population. Under this assumption that A_1 and A_2 have positive monotonic effects on the outcome, condition (1) would suffice to draw the conclusion of the presence of a sufficient-cause interaction[2,6] and the contrast $p_{11} - p_{10} - p_{01} + p_{00}$ then gives a lower bound on the prevalence of individuals with a sufficient-cause interaction. Testing condition (2) would be necessary without such monotonicity assumptions.[1,2] It should be noted that conditions (1) and (2) are, under the assumptions given above, sufficient but not necessary conditions for a sufficient-cause interaction. A sufficient-cause interaction may be present even if these conditions are not satisfied.

VanderWeele[15,16] discussed empirical tests for an even stronger notion of interaction. In the genetics literature, when gene-gene interactions are in view, "compositional epistasis"[17,18] is said to be present if there is some person such that $D_{11} = 1$ and $D_{10} = D_{01} = D_{00} = 0$. This means that for this person the outcome occurs if and only if both exposures are present. In such cases, the effect of one exposure will be masked if the other is absent i.e. the effect of the first genetic factor is 0, (and thus masked) unless the second genetic factor is present (and vice versa). Note that this is an even stronger notion of interaction than that of a sufficient-cause interaction, in that we are now requiring that $D_{00} = 0$. We might refer to this as an "epistatic interaction"[16] or a "singular interaction".[4] Although it is in general a stronger notion of interaction, if at least one of the two exposures has a positive monotonic effect on the outcome, then the notions of a "sufficient cause" interaction and a "singular/epistatic" interaction coincide.

VanderWeele[15,16] showed that if the effects of A_1 and A_2 on D were unconfounded, then

$$p_{11} - p_{10} - p_{01} - p_{00} > 0 \quad (3)$$

would imply an "epistatic interaction", i.e. the presence of compositional epistasis; the contrast $p_{11} - p_{10} - p_{01} - p_{00}$ in fact gives a lower bound on the prevalence of individuals that manifest such a singular/epistatic interaction. In contrast with condition (1), we subtract rather than add p_{00}. If at least one of the two exposures has a positive monotonic effect on the outcome, then we can test condition (2) to conclude the presence of a singular/epistatic

interaction. If both exposures have positive monotonic effects on the outcome, then we can use condition (1) to test for the presence of a singular/epistatic interaction.

Often epidemiologic data are obtained from case-control studies in which it is not possible to estimates the risks $p_{a_1a_2} = P(D = 1|A_1 = a_1, A_2 = a_2)$ directly. Instead, additive interaction is often assessed using a measure referred to as the "relative excess risk due to interaction" or RERI,[19] defined as:

$$RERI = RR_{11} - RR_{10} - RR_{01} + 1$$

where $RR_{a_1a_2} = p_{a_1a_2}/p_{00}$. In a case-control study where the odds ratio directly estimates a rate ratio or a relative risk (e.g. rare outcome or incidence density sampling), *RERI* can be estimated by $RERI \approx OR_{11} - OR_{10} - OR_{01} + 1$ where $OR_{a_1a_2} = \{p_{a_1a_2}/(1-p_{a_1a_2})\}/\{p_{00}/(1-p_{00})\}$. By dividing conditions (1)-(3) all by p_{00} we see that condition)1) could be tested by $RERI > 0$; condition (2) could be tested by $RERI > 1$; and condition (3) could be tested by $RERI > 2$. It is relatively straightforward to obtain confidence intervals for the relative excess risk due to interaction.[20–22]

Interference and Spillover Effects

We now consider a somewhat different setting in which only one exposure is in view but in which people are clustered such that the exposure that one person receives may affect the outcome of another person in the same cluster. For example, this might be the case for people living in households if the exposure were a vaccine. Suppose that in a particular study there are K households indexed by $i = 1, \ldots, K$ in which there are two people under study per household (e.g. husband and wife) indexed by $j = 1, 2$. Initially let us suppose that the two persons are distinguishable from one another (e.g. $j = 1$ denotes the wife and $j = 2$ denotes the husband). We let A_{ij} denote the exposure status for person j in household i. For example, for a vaccine we let $A_{ij} = 1$ denote that the person received the vaccination and $A_{ij} = 0$ that the person did not. We let Y_{ij} denote the infection status of person j in household i after some suitable follow-up. For reasons that will become clearer below we are now using Y rather than D to denote the outcome.

Now let $Y_{ij}(a_{i1}, a_{i2})$ denote the counterfactual outcome for person j in household i if the two people in that household i had (possibly contrary to fact) vaccine status of (a_{i1}, a_{i2}). For example, $Y_{i2}(1, 0)$ would denote what would have happened to person 2 if person 1 had received the vaccine and person 2 had not; $Y_{i1}(0, 0)$ denotes what would have happened to person 1 if neither had received the vaccine. Note that under this counterfactual or "potential outcomes" notation, the potential outcome for person 1, $Y_{i1}(a_{i1}, a_{i2})$, depends on the vaccine status of both persons. This allows for the possibility that the exposure status of one person affects the outcomes of another, sometimes referred to as interference or a spillover effect. Most literature in causal inference makes a "no-interference" assumption that one person's outcome does not depend on the exposure of other people. In the current context this would imply that $Y_{i1}(a_{i1}, a_{i2}) = Y_{i1}(a_{i1})$ and $Y_{i2}(a_{i1}, a_{i2}) = Y_{i2}(a_{i2})$ so that each person's outcome depends only on his or her own exposure status. We will allow for such interference here, but will assume that the vaccine status of people in one household do not affect the

outcomes of those in other households. This assumption is sometimes referred to as "partial interference".[8,11] This would not be an unreasonable assumption if the source population in the study were very large and a relatively small number of households were randomly selected for inclusion in the study. (As noted above, this use of "interference" is different from that in the infectious disease literature.)

If we allow for such interference between people within a household we can define various causal effects of interest beyond the overall effect of vaccinating versus not vaccinating households. For example we might consider a person effect or "direct effect"[11,23] of having one person (e.g. the wife) vaccinated while holding the vaccine status of the other person (e.g. the husband) constant. In counterfactual notation this would be $Y_{i1}(1, a_{i2}) - Y_{i1}(0, a_{i2})$; if we hold the husband's vaccine status to vaccinated this is $Y_{i1}(1, 1) - Y_{i1}(0, 1)$; if we hold the husband's vaccine status to unvaccinated this is $Y_{i1}(1, 0) - Y_{i1}(0, 0)$. Alternatively we could consider indirect or spillover effects such as the effect on the wife's outcome of having the husband vaccinated versus unvaccinated while holding the wife's vaccine status constant. In counterfactual notation this would be $Y_{i1}(a_{i1}, 1) - Y_{i1}(a_{i1}, 0)$; if we hold the wife's vaccine status to vaccinated this is $Y_{i1}(1, 1) - Y_{i1}(1, 0)$; if we hold the wife's vaccine status to unvaccinated this is $Y_{i1}(0, 1) - Y_{i1}(0, 0)$. We could also define analogous effects for the husband. In general we cannot hope to estimate these individual/direct and spillover/indirect effects for a particular household, but we may be able to estimate these effects on average, at least in various randomized trials in which both the husband's and the wife's exposures are randomized. We could do so by simply comparing the sample average of the wife's or husband's outcome across the various subgroups defined by the wife's and husband's vaccination status. Settings in which clusters have more than two people and with other randomization schemes are discussed elsewhere.[11,12,14]

If any of the various spillover/indirect effects in this setting are non-zero, then we would say that interference is present. For example if we found on average that

$$E\left[Y_{i1}\left(0,1\right) - Y_{i1}\left(0,0\right)\right] > 0$$

we would know that there was a spillover effect of the husband's vaccine on the wife's outcome. In the following section, by relating the notions of interference to causal interaction, we give new results on detecting specific forms of interference. These results will also shed light on the relationship between interference and causal interactions.

Tests for Specific Forms of Interference Using Causal Interactions

Suppose now that we are interested in trying to detect patterns of interference of a particular form. For example, we might be interested in whether there are any households such that the wife is not infected if and only if both the husband and the wife are vaccinated. Expressed in terms of counterfactuals, we would be asking whether there is some household *i* such that $Y_{i1}(1, 1) = 0$ but $Y_{i1}(1, 0) = Y_{i1}(0, 1) = Y_{i1}(0, 0) = 1$. This pattern is somewhat analogous to the epistatic/singular interaction considered above. In fact, by redefining our outcome, we can test for it empirically. For each household *i*, define $D_i(a_{i1}, a_{i2}) = 1 - Y_{i1}(a_{i1}, a_{i2})$; i.e. $D_i(a_{i1}, a_{i2})$ is an indicator that the wife is not infected if the wife and husband receive

vaccines corresponding to a_{i1} and a_{i2}, respectively. Suppose that both the wife's and husband's vaccine status were randomized. If we then let $p_{a_1a_2} = P(D = 1|A_1 = a_1, A_2 = a_2) = P(Y_1 = 0|A_1 = a_1, A_2 = a_2)$, then by using the tests for causal interaction we would have that if

$$p_{11} - p_{10} - p_{01} - p_{00} > 0$$

there must be some households such that the wife is not infected if and only if both the husband and the wife are vaccinated. In fact, the contrast $p_{11} - p_{10} - p_{01} - p_{00}$ will be a lower bound on the prevalence of such households. This is now a conclusion concerning a much more specific form of interference than simply the presence of some spillover effect, as in the previous section. By redefining $D_i(a_{i1}, a_{i2}) = 1 - Y_{i2}(a_{i1}, a_{i2})$, we could test for similar patterns of interference for the husband. Similarly, by yet other alternative definitions for D_i we could attempt to detect yet other forms of potential interference. If we define $D_i(a_{i1}, a_{i2}) = [1 - Y_{i1}(a_{i1}, a_{i2})] \times [1 - Y_{i2}(a_{i1}, a_{i2})]$ and find, for $p_{a_1a_2} = P(D = 1|A_1 = a_1, A_2 = a_2)$ with D so defined, that $p_{11} - p_{10} - p_{01} - p_{00} > 0$, then we could conclude that there were households such that the husband and the wife both remain uninfected if and only if both receive the vaccine. By defining D_i as other combinations of the wife's and husband's outcomes, other forms of interference or response patterns could potentially be tested for. As with causal interactions, however, so also here with tests for specific forms of interference, the conditions tested are sufficient for the specific form of interference in question, but they are not necessary. Such forms of interference might be present even if the condition on the probabilities is not satisfied.

However, as was also the case of "causal interactions", so too here, when testing for various forms of interference, monotonicity conditions will allow one to test weaker conditions. Consider again trying to test for whether there are any households such that the wife is not infected if and only if both the husband and the wife are vaccinated. Suppose that we thought the wife's vaccine would never cause the wife to be infected, so that monotonicity held for A_{i1} (i.e. if $D_i(a_{i1}, a_{i2}) = 1 - Y_{i1}(a_{i1}, a_{i2})$ denotes the absence of the wife being infected, then A_{i1} will have a positive monotonic effect on D_i). By the results on causal interaction it follows that for $p_{a_1a_2}$, defined as $p_{a_1a_2} = P(D = 1|A_1 = a_1, A_2 = a_2) = P(Y_1 = 0|A_1 = a_1, A_2 = a_2)$, if we found that

$$p_{11} - p_{10} - p_{01} > 0$$

then at least some households would have to be such that the wife is not infected if and only if both the husband and the wife are vaccinated. Similarly if we also thought that the vaccination of the husband would never, for any household, cause the wife to be infected, then to test for households with this specific interference pattern for the wife we could test

$$p_{11} - p_{10} - p_{01} + p_{00} > 0.$$

Note that in this case the monotonicity assumptions for the wife's and the husband's vaccine are somewhat different insofar as both pertain to the outcome status of the wife. These

monotonicity assumptions would be violated if the vaccine might itself cause the infection. But if, in a specific context, the monotonicity assumptions are thought reasonable, then these weaker conditions, $p_{11} - p_{10} - p_{01} > 0$ or $p_{11} - p_{10} - p_{01} + p_{00} > 0$, could be tested; without these monotonicity assumptions the more stringent condition $p_{11} - p_{10} - p_{01} - p_{00} > 0$ would be tested. Other forms of interference corresponding to the outcome occurring for one person but not the other could be formed by simply redefining D_i as the relevant function of Y_{i1} and Y_{i2}. It should be noted that when one is testing for other forms of interference or other response patterns - such as whether both remain uninfected if and only if both receive the vaccine or whether at least one remains uninfected if and only if both receive the vaccine - the monotonicity assumptions that are considered will change because the definition of the outcome $D_i(a_{i1}, a_{i2})$ changes.

We have thus far been considering forms of interference analogous to those of epistatic/singular interactions, but similar results pertain to patterns of interference analogous to sufficient-cause interactions. For example, if we were interested in whether there are households such that the wife would be uninfected if both received the vaccine but would be infected if only one or the other of the spouses received the vaccine then this would be analogous to a sufficient-cause interaction. We then could, for example, test for such a form of interference, without making monotonicity assumptions, using the condition for sufficient-cause interaction without monotonicity, namely, $p_{11} - p_{10} - p_{01} > 0$. Note that for these forms of interference corresponding to sufficient-cause interactions, conclusions are not being drawn regarding what occurs if both persons are unvaccinated as they were with the analog of the epistatic/singular interactions. Further variations on these ideas are also possible insofar as we could draw conclusions about what sorts of outcomes might occur if and only if one person is vaccinated and the other unvaccinated by recoding exposure status and not simply outcomes. Such forms of interference would then be somewhat analogous to antagonism[24] in the context of causal interactions.

The discussion to this point has assumed that we can distinguish between people in the household (e.g. the husband and the wife), i.e. that the subscript labelings $j = 1$ and $j = 2$ are meaningful. This may not always be the case; for example, we may have data only on the number vaccinated in each household and the number who have the outcome in each household. Alternatively, we may be considering siblings such that there is no clear classification of $j = 1$ and $j = 2$. Suppose once again that both exposures are randomized. We could then define D_i as, say, that both people have the outcome (or, alternatively, don't have the outcome; or that at least one has the outcome; etc.). If in each cluster we arbitrary select one person as $j = 1$ and the other as $j = 2$ (as will be seen below it will not matter which is which) then we could define $D_i(a_{i1}, a_{i2})$ as, for example, both persons having the outcome if we set $A_{i1} = a_{i1}$ and $A_{i2} = a_{i2}$ and let $p_{a_1a_2} = P(D = 1|A_1 = a_1, A_2 = a_2)$. Now let A_i denote the number who received the exposure in cluster or household i (i.e. $A_i = 0$, 1 or 2) and let $p_a = P(D = 1|A = a)$; furthermore note if the exposures received by both people are randomized with the same probability of receiving the exposure, then

$$p_1 = P(D=1|A=1) = \frac{1}{2}P(D=1|A_1=1, A_2=0) + \frac{1}{2}P(D=1|A_1=0, A_2=1) = \frac{1}{2}(p_{10}+p_{01}).$$

By the arguments above, we could test whether there are clusters where both persons would

have the outcome if and only if both were exposed by testing $p_{11} - p_{10} - p_{01} - p_{00} > 0$ which since $p_{11}=p_2$, $\frac{1}{2}(p_{10}+p_{01})=p_1$, $p_{00}=p_0$ is equivalent to

$$p_2 - 2p_1 - p_0 > 0$$

Under the assumption that the exposure for at least one person in each cluster had a positive monotonic effect on the outcome D_i we could, by similar arguments, instead test

$$p_2 - 2p_1 > 0$$

to conclude that this form of interference was present and if the exposure for both persons in each cluster had a positive monotonic effect on the outcome D_i we could test

$$p_2 - 2p_1 + p_0 > 0.$$

As before, by recoding the outcome D_i or the exposures, we could also form tests for other forms of interference.

We note that these tests apply to settings in which there are more than two people per cluster (e.g. there may be varying number of children within each household), provided that the randomization of the exposure for persons $j = 1$ and $j = 2$ does not depend on the exposure status of the other people in the household. In the Appendix we generalize these results further, and consider settings in which more than two people are potentially randomized to the exposure or in which the exposure may have more than two levels. We in fact show that the entire theory of causal interaction for *n*-way interactions between exposures[4] and also for multivalued exposures[3,15,16] maps onto tests for specific forms of interference.

Illustration

Consider a hypothetical vaccine trial with two persons per household as described above, in which each of the two persons is randomized to receiving the vaccine or a placebo with probability 1/2. Let $q_{uv}^{rs}=P(Y_{i1}=r, Y_{i2}=s|A_{i1}=u, A_{i2}=v)$ and suppose that the results from the vaccine trial, consisting of the infection status probabilities, q_{uv}^{rs}, are as given in the Table. We will assume a very large trial and, for the purposes of the illustration, ignore sampling variability, i.e. assume that the Table represents the true infection-status probabilities.

Suppose we are interested in whether there are households such that both persons become infected if and only if neither were vaccinated. Without making any assumptions about monotonicity, by the arguments above, we could evaluate

$$q_{00}^{11} - q_{01}^{11} - q_{10}^{11} - q_{11}^{11} > 0.$$

In this case, we would have $q_{00}^{11} - q_{01}^{11} - q_{10}^{11} - q_{11}^{11} = 0.12 - 0.05 - 0.04 - 0.01 = 0.02 > 0$ which would indicate that there were households such that both persons became infected if and only if neither were vaccinated.

Suppose now instead we were interested in examining whether there are households such that neither person would be infected if and only if both were vaccinated. By the arguments above, we would be able to draw this conclusion without monotonicity assumptions if $q_{11}^{00} - q_{01}^{00} - q_{10}^{00} - q_{00}^{00} > 0$; or, under the assumption that for at least one person, administering the vaccine was never causative of infection for either, if $q_{11}^{00} - q_{01}^{00} - q_{10}^{00} > 0$; or, under the assumption that for both people administering the vaccine was never causative of infection for either, if $q_{11}^{00} - q_{01}^{00} - q_{10}^{00} + q_{00}^{00} > 0$. In this case, neither of the first two empirical conditions hold; however, for the third we have that $q_{11}^{00} - q_{01}^{00} - q_{10}^{00} + q_{00}^{00} = 0.96 - 0.81 - 0.83 + 0.69 > 0$. Thus if we were willing to assume that for both people receiving the vaccine was never causative of infection for either, we could draw the conclusion that there were households such that neither person would be infected if and only if both were vaccinated.

Discussion

We have considered the correspondence between various forms of causal interaction and specific forms of interference in which the exposure of one person may affect the outcome of another. We have seen how empirical tests for causal interactions[1–4,15,16] can be applied almost immediately to detect various forms of interference. In the context of interference, by defining a new outcome as some function of the outcomes of the various people in each cluster, such tests can be adapted to attempt to detect a wide range of different forms of interference. The approach applies to settings in which the people within a cluster are or are not distinguishable from one another. Similar ideas also extend to correspondences between *n*-way causal interactions for multiple exposures and interference patterns amongst *n* persons per cluster and also extend further to multivalued exposures. The entire theory of causal interaction has an analog within the context of detecting various forms of interference.

The results here have generally been cast within the context of randomized exposures. However, all of the results are still applicable if randomization is conditional on strata of cluster covariates C_i and analysis is done conditional on C_i (or, somewhat more generally, if the effects of the exposure status for a cluster is unconfounded conditional on cluster covariates $C_i$and analysis is again done conditional on C_i). However, modeling difficulties arise in cases in which C_i contains numerous confounding variables, or continuous covariates, or distinct covariate values for each person in a cluster. The conceptual issues concerning confounding control become somewhat more subtle in the context of interference than in the simple no-interference setting;[12] further development of these ideas will be pursued in future work.

The work here is also subject to the limitation of the assumption about “partial interference”, i.e. that there is no interference between clusters. In the case of a vaccine trial, where a few

households are randomly sampled from a large city and the cluster is treated as the household, this assumption is perhaps not unreasonable - but it is unlikely to hold exactly. In other settings, however, additional complications will likely arise. The present work (and the general approach taken in statistical work on interference to date[8–14]) has assumed the existence of well-defined clusters. In many settings this assumption too will not be reasonable. Recent work on social networks in effect relaxes this assumption of well-defined clusters, but the use of social networks to draw inferences about causality is still controversial.[25–29] In settings in which clusters are not well defined or where partial interference is clearly an unreasonable assumption, traditional transmission system models in the infectious disease literature constitute a much more adequate modeling approach.[30,31]

Before closing, we would like to note that the train of thought that led to the exploration of the correspondence between interaction and interference had a somewhat unlikely origin. Independent observations amongst co-authors on the formal similarity between empirical tests for causal interaction and one of the forms of Bell's inequality[32] in quantum mechanics (concerning the interference of the measurement of the spin of one electron on the other) had two consequences. First, it led to the insight that theory for causal interaction could provide a very simple proof of Bell's inequality.[33] Second, the formal similarity led to questions about the possible implications for epidemiology of this formal relation, and whether such formal relations might have implications for the phenomenon known as interference or spillover effects. Further reflection led to the realization that, in fact, the entire theory of causal interactions had analogs in the context of interference.

One further point concerning a resemblance in language is also worth noting. In the econometrics literature the phenomenon of "interference" or "spillover effects" is sometimes referred to as "social interactions".[34–36] This phenomenon arises in the study of neighborhoods, classrooms, judicial panels and elsewhere, whenever people interact with one another in such a way that the outcome of one person depends on the exposure of another. The results of this paper show that such "social interactions" may in some instances also be cases of "causal interaction." The use of the word "interaction" to describe both the joint action of two exposures and also the phenomenon of interference is, however, not simply a linguistic coincidence, nor merely indicative of a conceptual similarity, but instead (as we have now shown) connotes an extensive formal correspondence between various forms of causal interaction on the one hand and specific forms of interference on the other. The theory of causal interactions[1–4,15,16] turns out to provide a theoretical framework and analytical tool by which to reason about interference. Interference arises not simply in the context of vaccines but in the study of neighborhoods, classrooms, and judicial panels, as well as in various forms of economic behavior. The tests we have described here may be of use in detecting specific forms of interference in all of these settings. Here we have illustrated the results in discerning what can be learned about spillover effects from certain forms of vaccine trials. The broader implications of the correspondence we have described remains to be worked out.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Financial support: National Institutes of Health grants ES017876 and HD060696.

Appendix. Generalizations to Many People per Cluster and to Multi-Valued Exposures

Here we show that the relationships between causal interaction and specific forms of interference extend to settings with multiple persons per cluster. For causal interactions with three binary exposures of interest, let $D_{a_1a_2a_3}$ denote the counterfactual outcome for a person for some binary variable D if A_1, A_2, and A_3 had been set, possibly contrary to fact, to a_1, a_2 and a_3 respectively. We say that there is a 3-way sufficient-cause interaction between A_1, A_2, and A_3 if there is a person such that $D_{111} = 1$ and $D_{110} = D_{101} = D_{011} = 0$. VanderWeele and Robins[2] noted that if we let $p_{a_1a_2a_3} = E(D|A_1 = a_1, A_2 = a_2, A_3 = a_3)$, then if the effects of A_1, A_2 and A_3 on D are unconfounded, and if

$$p_{111} - p_{110} - p_{101} - p_{011} > 0 \quad (4)$$

then a sufficient cause interaction must be present between A_1, A_2 and A_3. If the effects of A_1, A_2 and A_3 on D are positive monotonic and unconfounded, then any of the following three conditions imply the presence of a three-way sufficient cause interaction[2,4]:

$$\begin{aligned} p_{111} - p_{110} - p_{101} - p_{011} + p_{100} + p_{010} &> 0 \\ p_{111} - p_{110} - p_{101} - p_{011} + p_{100} + p_{001} &> 0 \\ p_{111} - p_{110} - p_{101} - p_{011} + p_{010} + p_{001} &> 0. \end{aligned} \quad (5)$$

If just two of the exposures, say A_1 and A_2, have positive monotonic effects on the outcome, then the following condition suffices:

$$p_{111} - p_{110} - p_{101} - p_{011} + p_{001} > 0. \quad (6)$$

Consider now instead the context of interference in which there are three people per cluster and A_{i1}, A_{i2} and A_{i3} denote the exposures received by persons 1, 2 and 3, respectively in cluster i, and let $Y_{ij}(a_{i1}, a_{i2}, a_{i3})$ denote the counterfactual outcome for person j in cluster i if A_{i1}, A_{i2} and A_{i3} had been set to a_{i1}, a_{i2}, a_{i3}, respectively. We could then define $D_i(a_{i1}, a_{i2}, a_{i3})$ as some function of $Y_{i1}(a_{i1}, a_{i2}, a_{i3})$, $Y_{i2}(a_{i1}, a_{i2}, a_{i3})$, $Y_{i3}(a_{i1}, a_{i2}, a_{i3})$. For example, $D_i(a_{i1}, a_{i2}, a_{i3})$ might simply be whether person 1 has the outcome when $A_{i1} = a_{i1}$, $A_{i2} = a_{i2}$, $A_{i3} = a_{i3}$ i.e. $D_i(a_{i1}, a_{i2}, a_{i3}) = Y_{i1}(a_{i1}, a_{i2}, a_{i3})$. Alternatively $D_i(a_{i1}, a_{i2}, a_{i3})$ could be taken as an indicator of all three persons having the outcome or at least two having the outcome, or the first and the second person but not the third, etc. In any case, regardless how $D_i(a_{i1}, a_{i2}, a_{i3})$ is defined, if we let $p_{a_1a_2a_3} = P(D = 1|A_1 = a_1, A_2 = a_2, A_3 = a_3)$, we could test whether there are clusters such that $D_i(a_{i1}, a_{i2}, a_{i3}) = 1$ whenever all three people have the exposure present, but not when just two of the three have the exposure, by testing condition (4). For example, if we let $p_{a_1a_2a_3} = P(Y_1 = 1|A_1 = a_1, A_2 = a_2, A_3 = a_3)$ and found that inequality (4) held then we could conclude that there were clusters in which the first person would have the outcome if all three people had the exposure but not if just two of three did.

As before, once $D_i(a_{i1}, a_{i2}, a_{i3})$ is defined, if we thought that the exposure for all three people had a positive monotonic effects on the outcome D_i, as defined, then we could test condition (5) instead of condition (4); if it were thought that the exposures of persons 1 and 2 had positive monotonic effects on the outcome, then one could test condition (6) instead of condition (4) to test for the particular form of interference or pattern of outcome responses under question. As with two people per cluster, when there are three persons per cluster but they are indistinguishable from one another if A_i denotes the number who received the exposure in cluster or household i (i.e. $A_i = 0, 1, 2$ or 3) and we let $p_a = P(D = 1|A = a)$ then if the exposure is randomized for each person with the same probability then conditions (4), (5) and (6) can be rewritten respectively as:

$$p_3 - 3p_2 > 0$$

for condition (4);

$$p_3 - 3p_2 + 2p_1 > 0$$

for condition (5) and

$$p_3 - 3p_2 + p_1 > 0$$

for condition (6). Analogs of the epistatic/singular interaction for three binary exposures are given in VanderWeele and Richardson[4] and could likewise be applied to test for corresponding forms of interference. VanderWeele and Richardson[4] also consider both sufficient cause and epistatic/singular n-way interactions, and these could likewise be used to test for various forms of interference when there are n persons per cluster. The entire theory of causal interaction for n-way exposures maps onto tests for specific forms of interference among clusters with n persons.

Likewise the theory for causal interactions for two multi-valued exposures maps onto tests for specific forms interference for multivalued exposures between two people in a cluster. Theory has been developed for both sufficient-cause interaction for two exposures each with three levels[3] and for epistatic/singular interactions for two exposures each with three levels[15,16]. For example, for two exposures, A_1 and A_2, each with three levels, with $D_{a_1a_2}$ denoting the counterfactual outcome setting A_1 to a_1 and A_2 to a_2 and $p_{a_1a_2} = P(D = 1|A_1 = a_1, A_2 = a_2)$ then if the effects of A_1 and A_2 on D are unconfounded there will be people such that $D_{a_1a_2} = 1$ if and only if $a_1 = a_2 = 2$ if

$$p_{22} - p_{21} - p_{20} - p_{12} - p_{11} - p_{10} - p_{02} - p_{01} - p_{00} > 0.$$

Considerably weaker conditions can be tested under monotonicity assumptions or if sufficient-cause interactions, rather than epistatic/singular interactions, are in view. However, as before, these various empirical tests all have analogs for detecting various forms of interference between clusters with two people when the exposures have three levels (e.g. no vaccine, low-dose vaccine, high-dose vaccine). Here again, $Y_{ij}(a_{i1}, a_{i2})$ would

denote the counterfactual outcome for person j in cluster i if A_{i1} and A_{i2} had been set to a_{i1} and a_{i2} respectively, $D_{ij}(a_{i1}, a_{i2},)$ could be defined as some function of $Y_{i1}(a_{i1}, a_{i2})$ and $Y_{i2}(a_{i1}, a_{i2})$, and the the tests for either sufficient-cause or epistatic/singular interactions for categorical or ordinal variables could once again essentially be applied directly either with or without monotonicity assumptions. Conditions in which one exposure has two levels and the other three are also available.[3,15,16] Once we consider exposures with three or more levels, the number of possible forms of causal interaction (and forms of interference) and conditions that may be tested under varying monotonicity assumptions increases considerably; the interested reader can find these tests for causal interactions elsewhere.[3,15,16] Analogous extensions are conceivable for multivalued exposures in clusters with an arbitrary number of people. We see that the close correspondence between causal interaction and forms of interference thus extends considerably further: the entire theory of causal interaction for n-way interactions between exposures[4] and also for multivalued exposures[3,15,16] maps onto tests for specific forms interference.

References

1. VanderWeele TJ, Robins JM. The identification of synergism in the sufficient-component cause framework. Epidemiol. 2007; 18(3):329–339.
2. VanderWeele TJ, Robins JM. Empirical and counterfactual conditions for sufficient cause interactions. Biometrika. 2008; 95(1):49–61.
3. VanderWeele TJ. Sufficient cause interactions for categorical and ordinal exposures with three levels. Biometrika. 2010; 97:647–659. [PubMed: 22822251]
4. VanderWeele TJ, Richardson TS. General theory for interactions in sufficient cause models with dichotomous exposures. Annals of Statistics. conditionally accepted.
5. Rothman KJ. Causes. Am. J. Epidemiol. 1976; 104(6):587–592. [PubMed: 998606]
6. Rothman, KJ.; Greenland, S.; Lash, TL. Modern Epidemiology. 3rd edition. Lippincott Williams and Wilkins; Philadelphia: 2008.
7. VanderWeele TJ. Sufficient cause interactions and statistical interactions. Epidemiol. 2009; 20(1): 6–13.
8. Sobel ME. What do randomized studies of housing mobility demonstrate?: Causal inference in the face of interference. Journal of the American Statistical Association. 2006; 101:1398–1407.
9. Hong G, Raudenbush SW. Evaluating kindergarten retention policy: A case study of causal inference for multilevel observational data. Journal of the American Statistical Association. 2006; 101:901–910.
10. Rosenbaum PR. Interference between units in randomized experiments. Journal of the American Statistical Association. 2007; 102:191–200.
11. Hudgens MG, Halloran ME. Towards causal inference with interference. Journal of the American Statistical Association. 2008; 103:832–842. [PubMed: 19081744]
12. Tchetgen Tchetgen EJ, VanderWeele TJ. Estimation of causal effects in the presence of interference. Statistical Methods in Medical Research - Special Issue on Causal Inference. in press.
13. VanderWeele TJ, Tchetgen Tchetgen EJ. Bounding the infectiousness effect in vaccine trials. Epidemiology. 22:686–693. [PubMed: 21753730]
14. VanderWeele TJ, Tchetgen Tchetgen EJ. Effect partitioning under interference for two-stage randomized experiments. Statistics and Probability Letters. 2011; 81:861–869. [PubMed: 21532912]
15. VanderWeele TJ. Empirical tests for compositional epistasis. Nature Reviews Genetics. 2010; 11:166.
16. VanderWeele TJ. Epistatic interactions. Statistical Applications in Genetics and Molecular Biology. 2010; 9:1–22. Article 1.

17. Phillips PC. Epistasis – the essential role of gene interactions in the structure and evolution of genetic systems. Nature Reviews Genetics. 2008; 9:855–867.
18. Cordell HJ. Detecting gene-gene interaction that underlie human diseases. Nature Reviews Genetics. 2009; 10:392–404.
19. Rothman, KJ. Modern Epidemiology. 1st ed.. Little, Brown and Company; Boston, MA: 1986.
20. Hosmer DW, Lemeshow S. Confidence interval estimation of interaction. Epidemiology. 1992; 3:452–56. [PubMed: 1391139]
21. Richardson DB, Kaufman JS. Estimation of the relative excess risk due to interaction and associated confidence bounds. Am J Epidemiol. 2009; 169:756–60. [PubMed: 19211620]
22. VanderWeele TJ, Vansteelandt S. A weighting approach to causal effects and additive interaction in case-control studies: marginal structural linear odds models. American Journal of Epidemiology. 2011; 174:1197–1203. [PubMed: 22058231]
23. Halloran ME, Struchiner CJ. Causal inference for infectious diseases. Epidemiology. 1995; 6:142–51. [PubMed: 7742400]
24. VanderWeele TJ, Knol MJ. Remarks on antagonism. American Journal of Epidemiology. 2011; 173:1140–1147. [PubMed: 21490044]
25. Carrington, PJ.; Scott, J.; Wasserman, S. Model and methods in social network analysis. Cambridge University Press; New York: 2005.
26. Christakis NA, Fowler JH. The spread of obesity in a large social network over 32 years. New England Journal of Medicine. 2007; 357:370–379. [PubMed: 17652652]
27. Shalizi CR, Thomas AC. Homophily and contagion are generically confounded in observational social network studies. Sociological Methods and Research. 2011; 40:211–239. [PubMed: 22523436]
28. VanderWeele TJ. Sensitivity analysis for contagion effects in social networks. Sociological Methods and Research. 2011; 40:240–255. [PubMed: 25580037]
29. Lyons R. The spread of evidence-poor medicine via flawed social network analysis. Statistics, Politics, and Policy. 2011; 2 Article 2.
30. Diekmann, O.; Heesterbeek, JAP. Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation. Wiley; New York: 2000.
31. Riley S. Large-scale spatial-transmission models of infectious disease. Science. 2007; 316:1298–1301. [PubMed: 17540894]
32. Bell JS. On the Einstein Podolsky Rosen paradox. Physics. 1964; 1:195–200.
33. Robins, JM.; VanderWeele, TJ.; Gill, RD. A proof of Bell's inequality in quantum mechanics using causal interactions. Collection of Biostatistics Research Archive; Preprint Series, Article 83. http://biostats.bepress.com/cobraffps/art83
34. Graham B. Identifying social interactions through conditional variance restrictions. Econometrica. 2008; 76:643–660.
35. Manski CF. Economic analysis of social interactions. Journal of Economic Perspectives. 2000; 14:115–136.
36. Manski, CF. Identification of treatment response with social interactions. Northwestern University Technical Report.

Table

Probabilities of infection (Y_{i1}, Y_{i2}) by vaccination status (A_{i1}, A_{i2})

	$Y_{i1} = 0, Y_{i2} = 0$	$Y_{i1} = 0, Y_{i2} = 1$	$Y_{i1} = 1, Y_{i2} = 0$	$Y_{i1} = 1, Y_{i2} = 1$
$A_{i1} = 0, A_{i2} = 0$	0.69	0.10	0.09	0.12
$A_{i1} = 0, A_{i2} = 1$	0.81	0.05	0.09	0.05
$A_{i1} = 1, A_{i2} = 0$	0.83	0.10	0.03	0.04
$A_{i1} = 1, A_{i2} = 1$	0.96	0.02	0.01	0.01