

# Dependent Happenings: a Recent Methodological Review

M. Elizabeth Halloran<sup>1,2</sup> · Michael G. Hudgens<sup>3</sup>

Published online: 28 July 2016  
© Springer International Publishing AG 2016

**Abstract** One hundred years ago, Sir Ronald Ross published his treatise on a general theory of happenings. Dependent happenings are those in which the frequency depends on the number already affected. When there is dependency of events, interventions can have different types of effects. Interventions such as vaccination can have direct protective effects for the person receiving the treatment, as well as indirect/spillover effects for others in the population. Causal inference is a framework for carefully defining the causal effect of a treatment, exposure, or policy, and then determining conditions under which such effects can be estimated from the observed data. We consider here scenarios in which the potential outcomes of an individual can depend on the treatment of other individuals in the population, known as causal inference with interference. Much of the research so far has assumed the population is divided into groups or clusters, and individuals can interfere with others within their clusters but not across clusters. Recent developments have assumed more general forms of interference. We review some of the different types of effects that have been defined for dependent happenings, particularly using the methods of causal inference with interference. Many of the methods are applicable across

disciplines, such as infectious diseases, social sciences, and economics.

**Keywords** Causal inference · Counterfactual · Dependent happenings · Experimental design · Herd immunity · Indirect effects · Networks · Peer influence effects · Potential outcome · Spillover effects · SUTVA

## Introduction

One hundred years ago, Sir Ronald Ross [1] published his treatise on a general theory of happenings. The happening element corresponds in modern parlance to an occurrence or event. Ross differentiated independent from dependent happenings. Independent happenings were “those in which the frequency of the happening is *independent* of the number of individuals already affected” ([1], p. 211). Dependent happenings were those in which the frequency depends on the number already affected. To the first class belong such happenings as noninfectious diseases and accidents. To the second class belong “infectious diseases, membership of societies and sects with propagandas, trade-unions, political parties, etc., due to propagation from within, that is, individual to individual” ([1], p. 211). Ross had been awarded the second Nobel prize in Medicine in 1902 for elucidating the role of mosquito vectors in the transmission cycle of malaria. He was also an amateur mathematician and developed the original malaria models to quantify the effects of interventions [2, 3]. Ross had published an addendum in the second edition of his book on the prevention of malaria on a preliminary general theory of happenings in 1911.

A principled approach to evaluating the effects of interventions that has fostered clarity in thinking about both experimental and observational studies is causal inference [4–7].

---

This article is part of the Topical Collection on *Epidemiologic Methods*

✉ Michael G. Hudgens  
mhudgens@email.unc.edu

<sup>1</sup> Center for Inference and Dynamics of Infectious Diseases, Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA

<sup>2</sup> Department of Biostatistics, School of Public Health, University of Washington, Seattle, WA, USA

<sup>3</sup> Department of Biostatistics, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

Causal inference is a framework for carefully defining the causal effects of a treatment, exposure, or policy, then determining conditions under which such effects can be estimated from the observed data. In causal inference, each individual is assumed to have a potential outcome that would occur under each of the different possible treatments. The outcome is potential in that it only occurs if, in fact, a particular treatment is given. Usually the treatment assignment of one person is assumed to not affect the potential outcomes of another person, called the assumption of *no interference* by Cox [8]. The assumption of no interference is a component of the Stable Unit Treatment Value Assumption (SUTVA) [9]. The assumption of no interference is also called the individualistic treatment response [10]. Under dependent happenings, interference will likely be present. In the presence of interference, treatments can have direct effects on the person receiving the treatment as well as indirect effects on people not receiving treatment. An individual may have many different potential outcomes depending on the treatment status of the others.

Several branches of research have their roots in the work by Ross. One is the broad field of infectious disease transmission models, including modeling the effects of interventions in populations, which Ross called a priori pathometry. Another is the large and growing literature on estimating parameters for these transmission models. Another area is methods to evaluate interventions in populations based on data from field studies. In this review, our focus is on the latter, specifically, on drawing inference about effects of interventions in the presence of interference, that is, for dependent happenings.

Here, we consider examples largely in the epidemiological context, in particular infectious diseases and the effects of vaccines. However, much of the research on causal inference with interference is motivated by applications in the social sciences and economics, which also come under the general heading of dependent happenings in the sense of Ross. Many of the methods are applicable across disciplines.

## Partial Interference

Motivated by an interest in dependent happenings, Halloran and Struchiner [11, 12] defined direct, indirect, total, and overall effects in the presence of interference (Fig. 1). Consider a cluster (or group) of individuals under two scenarios. Under scenario A on the left, a certain portion of individuals in the cluster is vaccinated and the rest remains unvaccinated. Under scenario B on the right, no one in the cluster is vaccinated. The *direct effect* of vaccination under scenario A is defined by comparing the average outcome when an individual is vaccinated with the average outcome when an individual is not vaccinated. The *indirect effect* is defined as a contrast between the average outcome when an individual is not vaccinated under scenario A compared with scenario B. The *total effect*

is defined by comparing the average outcome when an individual is vaccinated under scenario A to the average outcome when an individual is not vaccinated in scenario B. In general, the total effect can be decomposed into a function of the direct and indirect effects [13, 14]. The *overall effect* is defined by the contrast in the average outcome in the entire cluster under scenario A compared to the average outcome of the entire cluster under scenario B.

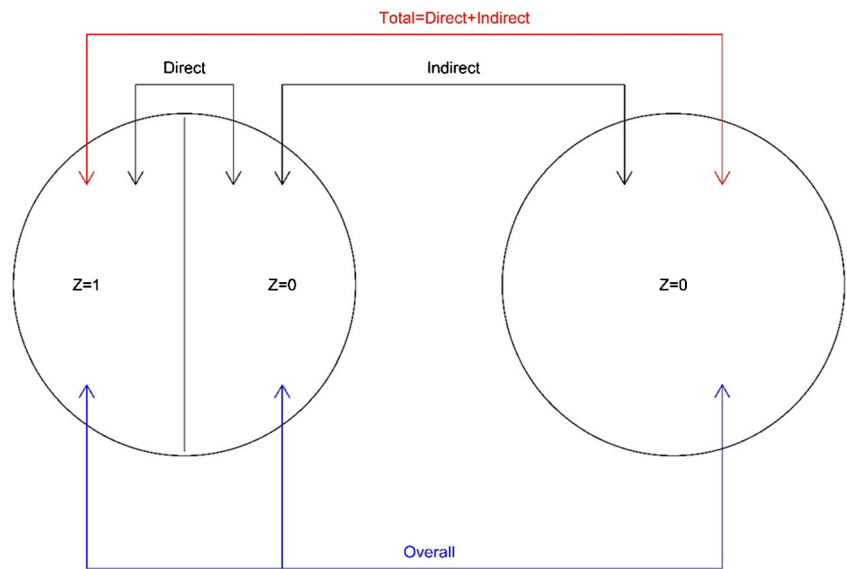
With vaccination, indirect effects are produced by an increase in herd immunity, the collective immunological status of the population. In other fields, indirect effects are sometimes called *spillover effects* or *peer influence effects* [15]. In mediation analysis, the terms direct, indirect, and overall effects have different meanings than used here [16].

The potential outcome approach to causal inference typically assumes no interference between individuals. Assuming no interference, if there is one treatment and a control, then an individual has two potential outcomes. The individual causal effect, which in general cannot be observed, is the difference between the two potential outcomes under treatment and control. The average causal effect is the average of the difference between the potential outcomes if everyone received treatment and if everyone received control. Drawing inference about treatment effects generally requires knowledge or modeling of the mechanism by which individuals select or are assigned treatment. Under an independent assignment mechanism, such as randomization, one can construct an unbiased estimator of the average causal effect from the observed average outcomes in the treatment and control groups [5].

The situation is more complicated under dependent happenings in the presence of interference. Halloran and Struchiner [12] proposed individual-level causal estimands 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 [9, 17]. However, with a binary outcome and one treatment and one control, if there are  $N$  people in a population, there are  $2^N$  possible treatment vectors and  $2^N$  possible potential outcomes. Causal inference in the presence of so many potential outcomes would seem difficult, if not impossible, without making additional assumptions.

One possible additional assumption which simplifies the problem considerably is to assume individuals can be partitioned into groups (or clusters) wherein there is no interference between individuals in different groups. Sobel [18] referred to this assumption as *partial interference*. As a motivating example, he considered interference in a housing mobility experiment in poor neighborhoods in five cities. Eligible residents were randomly assigned to receive one of two forms of relocation assistance or no assistance (control). It was plausible that many of the participants knew each other and might influence each other's behavior. Sobel assumed interference could occur within but not across sites, i.e., there was partial

**Fig. 1** Study designs for dependent happenings. A cluster is considered under two different scenarios. In the scenario A on the left, a certain portion of individuals in the cluster receive treatment,  $Z = 1$ , and the other portion of individuals receive control or nothing,  $Z = 0$ . In the scenario B on the right, everyone receives control or nothing. The direct, indirect, total, and overall effects of intervention are defined by the indicated contrasts (adapted from [11, 12])



interference. In the next sections, we consider approaches that assume partial interference.

### Two-Stage Randomization

Suppose there are two different vaccination strategies (or policies or programs) under consideration. The first strategy might be to vaccinate 60 % of the individuals in a population; the other strategy might be to vaccinate no one. A question of interest then is what are the direct, indirect, total, and overall effects of the 60 % vaccination strategy compared to the no vaccination strategy? Assuming partial interference, Hudgens and Halloran [13] defined group- and population-level causal estimands for direct, indirect, total, and overall causal effects of treatment under two different treatment allocations. To obtain unbiased estimators of the population-level causal estimands, Hudgens and Halloran [13] proposed a two-stage randomization scheme: the first stage at the group level and the second stage at the individual level within groups. For example, suppose there are 20 groups of individuals. At the first stage, we could randomize 10 groups to the 60 % vaccination strategy and the remaining groups to the no vaccination strategy. In the second stage within 10 groups, we would randomly assign 60 % of individuals to receive vaccine, and in the other 10 groups, no one would receive vaccine. Unbiased estimator’s of the direct, indirect, and total effects can be obtained by contrasts in average outcomes among vaccinated and unvaccinated individuals under the different vaccination strategies. Likewise, contrasts in average outcomes of all individuals under the two vaccination strategies yield unbiased estimator’s of the overall effect.

To draw inference about the four effects, variance estimators and confidence intervals are needed. Hudgens and Halloran [13] developed variance estimators under an additional assumption they call *stratified interference*, which

assumes the indirect treatment effects may depend only on the proportion of other individuals in the group which receive treatment. Tchetgen Tchetgen and Vander Weele [19•], Liu and Hudgens [20], and Rigdon and Hudgens [21] develop exact and asymptotic confidence intervals.

Baird et al. [22•] considered a similar two-stage randomized design in the context of economic experiments to measure indirect/spillover effects. They refer to the level of (treatment) coverage in a cluster as the *saturation level* and their study design as the *randomized saturation design*. Their causal estimands, the intention to treat effect, spillover on the non-treated effect, and total causal effect, are analogous to the total, indirect, and overall effects, respectively, defined above. Baird et al. also consider optimal design of two-stage randomized trials to detect the different effects. Based on these considerations, they designed an experiment in Malawi to assess whether cash transfers help adolescent girls improve schooling outcomes and also delay marriage and pregnancy. Their study design accounted for possible spillover effects to non-beneficiaries as well as the participating adolescent girls in the same communities.

### Partially Randomized and Observational Studies

Often studies are not randomized at two stages, but only at the individual level, the cluster level, or neither. For example, Sur et al. [23] describe a cluster-randomized study in India where 80 geographic clusters were randomized to receive either typhoid vaccine or control (hepatitis A) vaccine. A subset of individuals in these clusters chose to receive the study vaccine (typhoid or control). Similarly, Moulton et al. [24] designed a cluster-randomized study to estimate total, indirect, and overall effectiveness of pneumococcal vaccine in infants. The

design entailed randomizing clusters to receive the pneumococcal vaccine or a control vaccine, with individuals self-selecting to enroll or not in the trial. See Hayes and Moulton [25] for a general overview of cluster-randomized trials and Halloran [26] for the discussion of the related minicommunity design for assessing indirect effects of vaccination.

In studies which do not utilize two-stage randomization, the estimators described in the previous section would in general be biased or inconsistent due to potential confounding between treatment and the outcome. In the observational setting where the treatment assignment mechanism is not known and there is no interference, the propensity score, the probability an individual receives a treatment assignment based on a function of observed pretreatment covariates, is one method to adjust for confounding [27]. Recently propensity score methods have been extended to the setting where interference may be present. For example, Hong and Raudenbush [28] 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. They classified schools by whether they retained a high proportion or a low proportion of kindergartners. Interference within a school was assumed to be summarized in the dichotomous school treatment of high or low retention rate. The study was observational at two levels: schools were not randomized to have high or low retention, and students at risk to be retained were not randomized to be retained. Assuming partial interference, Hong and Raudenbush used a multilevel propensity stratification approach to draw inference about the effects of retention. The school propensity of adopting a high retention rate was estimated using pretreatment information. Then the child propensity of repetition in high-retention schools and the child propensity of repetition in low-retention schools were estimated also based on pretreatment information. Estimation of the causal effects used stratifications by the estimated school and child propensity scores.

Tchetgen Tchetgen and Vander Weele [19•] also used group-level propensity scores to develop inverse probability weighted (IPW) estimators of the direct, indirect, total, and overall causal effects in observational studies in the absence of two-stage randomization. These IPW estimators can be viewed as a generalization of the usual IPW estimator of the causal effect of a treatment in the absence of interference. Lundin and Karlsson [29] proposed similar IPW estimators of direct, indirect, and total effects under interference where treatment assignment is randomized only at the first stage, and in some groups, all individuals remain untreated (Fig. 1). They were motivated by the Positive Parenting Program, an Australian parenting support program tested in pre-schools in Uppsala, Sweden. Perez-Heydrich et al. [30] and Liu et al. [31] considered the asymptotic properties of different IPW estimators in the presence of partial interference.

## Example

To illustrate estimates of direct, indirect, total, and overall effects of vaccination, we present results of an analysis by Perez-Heydrich et al. [30]. They used the IPW estimators proposed by Tchetgen Tchetgen and Vander Weele [19•] to estimate the effects of cholera vaccination in a study in Matlab, Bangladesh. All children and women were individually randomized to either of two cholera vaccines or a placebo, though not all participated in the vaccine trial. Unvaccinated individuals included eligible non-participants and placebo recipients. Vaccinated individuals included recipients of either vaccine. This was an individually randomized study, so neighborhoods (groups) were defined by using a clustering algorithm on geo-referenced data of the location of individuals' households. Effect estimates were calculated based on the difference of the IPW average outcomes in the relevant groups (Fig. 2). Levels of vaccine coverage are denoted by  $\alpha$  and  $\alpha'$ .

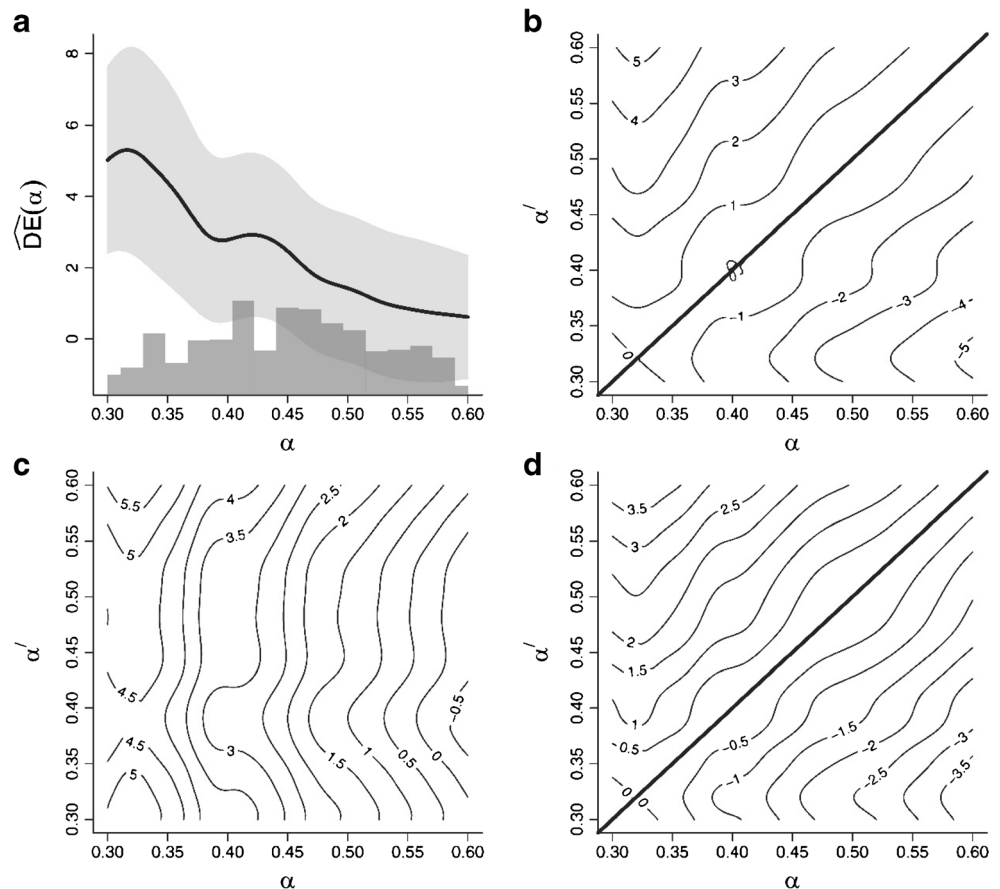
The direct effect estimates (Fig. 2a) generally decrease with increasing  $\alpha$ . The indirect effect estimates (Fig. 2b) comparing incidence of cholera among unvaccinated individuals at coverage levels  $\alpha$  and  $\alpha'$  tend to increase (in absolute value) as the difference in the coverage  $\alpha - \alpha'$  increases. The total effect contours (Fig. 2c) are roughly vertical, suggesting the estimated risk of cholera when vaccinated tends to be the same regardless of coverage. Note the total effect estimate of being vaccinated in a neighborhood with  $\alpha' = 60\%$  coverage compared to being unvaccinated in a neighborhood with  $\alpha = 32\%$  coverage is 5.9 (95% confidence interval 3.0, 8.8), an order of magnitude greater than the estimated direct effect of 0.6 at 60% coverage (Fig. 2a). Thus, taking indirect effects into account leads to a much different conclusion about the vaccine. The estimates of overall effects (Fig. 2d) exhibit a similar pattern to the indirect estimates. An R package "inference" is available for conducting this type of analysis [32].

## General Interference

Individuals do not necessarily congregate in identifiable groups wherein they interact only within those groups but not with individuals in other groups, as is assumed with partial interference. Recent methodological developments for causal inference in the presence of interference allow for more general forms of interference. In the general interference setting, each individual may be assumed to have a unique set of other individuals whose treatment might affect the outcome of the individual. This interference structure can be represented by a network.

Network theory, also called graph theory, provides a formal framework for describing interactions between individuals. A network is defined by a collection of  $N$  nodes or vertices, and the edges or ties connecting the nodes. In the context here, the

**Fig. 2** IPW estimates of **a** direct  $DE(\alpha)$ , **b** indirect  $IE(\alpha, \alpha')$ , **c** total  $TE(\alpha, \alpha')$ , and **d** overall  $OE(\alpha, \alpha')$  effects based on the cholera vaccine trial data. In **(a)** the gray region represents approximate pointwise 95 % confidence intervals. The histogram below depicts the distribution of observed neighborhood vaccine coverage. Units of the estimates are cases of cholera per 1000 individuals per year. (Reprinted with permission from Perez-Heydrich et al. [30])



nodes are individuals and the edges are connections to other individuals which may give rise to interference.

Interference and dependent happenings can be propagated through networks. Computational social science is a growing field, fueled by massive data sets on interactions that form social networks [33]. Companies such as Google and Facebook collect huge data sets on interactions and perform randomized experiments within these contexts. Wireless sensor network technology [34] and geographical positions from cell phone data [35] make electronic monitoring of close contacts and movement possible. Infectious diseases also propagate across networks on multiple scales [36].

Recent approaches to assessing treatment effects in the presence of general interference, some of which rely on network theory, include Rosenbaum [37], Toulis and Kao [15], Ugander et al. [38•], Eckles et al. [39], van der Laan [40•], Sofrygin and van der Laan [41], Aronow and Samii [42•], and Liu et al. [31] among others. These general interference methods typically assume one finite population of  $N$  individuals. For each individual, a set of other individuals is defined which may interfere with that individual. Different names have been used to describe these sets, including *interference sets* [31], *neighborhoods* [15, 38•], and *friends* [40•, 41] of the individuals. From the network perspective, the interference sets can be represented by an adjacency matrix. Typically,

the interference sets for each individual in the population are assumed known and fixed. It is also usually assumed that the number of individuals in these interference sets is smaller than  $N$  and that any indirect or spillover effects on an individual, if present, emanate from his/her known interference set.

Aside from these commonalities, there is considerable variability in recent approaches to causal inference with interference regarding how treatment/exposure effects (i.e., causal estimands) are defined and the mode of inference adopted. A comprehensive comparison is beyond this review, both in scope and technical detail. Below we highlight a few approaches recently proposed.

### Randomized Experiments

Aronow and Samii [42•] considered randomized experiments in the presence of arbitrary interference. Their approach depends on knowing the experimental design by which individuals are assigned treatment. They define an *exposure mapping* which relates the individual experimental treatment assignments to the exposures received by individuals depending on with whom they interact and other underlying features of the population. Their exposure mapping is equivalent to Manski’s [10] “effective treatments” function. A population is considered to have a number of possible exposure

conditions and causal estimands are defined as contrasts between average individual potential outcomes for two different exposure levels. The average potential outcome at any exposure level can be estimated because, by design, the probability of the exposure conditions is known for each individual. Aronow and Samii [42•] proposed IPW estimators for inference in this setting.

Ugander et al. [38•] and Eckles et al. [39] considered randomized experiments on networks where treatment of one individual could have indirect/spillover effects on neighboring individuals. Their causal estimand of interest is the average of the difference in the individual potential outcomes under two extreme assignments, one where every individual in a network receives treatment and one where no individuals receive treatment. If individuals were independently randomized to treatment and control, for individuals who share many edges with other individuals, the probability all of their neighbors would (i) all receive treatment or (ii) all receive control would be low. Thus, Ugander et al. considered a partition of the network into clusters of individuals and proposed randomizing some of the clusters to receive treatment and the remaining clusters to receive control. This is called *graph cluster randomization*. Note these clusters are not the same as the interference sets described above. Ugander et al. introduced the notion of *network exposure* wherein an individual is network exposed to treatment (control) if the individual's response under a particular assignment vector is the same as if everyone in the network had received the treatment (control). An unbiased estimator of the average treatment effect using inverse probability weighting was derived for any randomization design for which the network exposure probabilities can be explicitly computed.

Motivated by a randomized experiment by Facebook, Toulis and Kao [15] propose causal estimands for peer influence (indirect) effects describing interference in a social network. For each individual, a neighborhood is defined by the other individuals with whom he/she shares an edge. If at least one neighbor receives treatment, then the individual is considered *exposed to peer influence effects*. The potential outcome for each individual can depend on his treatment and that of his neighbors. Toulis and Kao defined two main causal estimands. The causal estimand for the *primary effect* is the average over the whole population of the difference in the individual outcomes if an individual received treatment versus received control when everyone else in the neighborhood received control. The main causal estimands for peer influence effects are defined by fixing the specific number of neighbors who receive treatment. For example, if  $k$  neighbors receive treatment, the  $k$ -level causal estimand for peer influence effects is averaged over individuals with at least  $k$  neighbors. Two estimation procedures are proposed: a frequentist model-based estimator assuming a certain sequential randomization design and known network and a Bayesian approach which accounts for uncertainty in the network topology.

We note there is a large, related literature on peer effects in social sciences dating back to Manski's seminal paper [43] on the "reflection" problem. Other recent developments of statistical methods for drawing inference about peer effects are described in An [44] and Vander Weele and An [45]. O'Malley and Onnela [46] give an overview, comparing approaches in social network analysis and network science.

### Observational Studies

van der Laan [40•] and Sofrygin and van der Laan [41] considered statistical inference about causal effects in the presence of general interference in the observational setting. They defined the population of interest to be a set of (possibly) dependent individuals and assumed only a single draw from the true data generating distribution is observed. That is, unlike traditional statistical inference, multiple independent and identically distributed (iid) replicates were not assumed. With partial interference, one might assume the groups are iid, permitting application of existing statistical theory (e.g., see [30]). However, with general interference observing iid replicates is generally not possible, such that standard large sample frequentist approaches do not apply. van der Laan [40•] and Sofrygin and van der Laan [41] derived the asymptotic properties of targeted maximum likelihood estimators in this setting, providing a method for valid statistical inference in the presence of general interference. See also Ogburn and Vander Weele [47] for related work allowing for general interference in the observational setting.

### Infectiousness and Contagion Effects

It is often of interest in causal inference to understand the mechanism or pathway through which an exposure or treatment has an effect on the outcome of interest. Below we describe two possible causal mechanisms of interference: the infectiousness and contagion effects.

The vaccine effect on infectiousness is the reduction in transmissibility from a vaccinated versus unvaccinated infected person during a contact with a susceptible person. In a study in Niakhar, Senegal, for example, Préziosi and Halloran [48] estimated the relative reduction in infectiousness of a vaccinated pertussis case compared to an unvaccinated case to be 67 % (95 % confidence interval 29, 86). Even if vaccination is randomized, the infectiousness effect is measured only in people who become infected, a post-randomization, variable, so naive estimates comparing infected vaccinees and infected unvaccinated individuals could be subject to selection bias [49]. Combining causal inference with

interference with principal stratification [50], Vander Weele and Tchetgen Tchetgen [51] and Halloran and Hudgens [52] proposed causal estimands of the infectiousness effect in households of size two, assuming interference occurred only within households, that is partial interference. Though this causal effect is not identifiable without further assumptions, bounds can be identified [51–53] and sensitivity analyses performed [54].

Widespread vaccination in a population can produce indirect effects either by preventing individuals from becoming infected or by reducing the ability of infected vaccinated individuals to transmit to others. Vander Weele et al. [16] differentiated these two components, calling the first the *contagion effect* and the second the *infectiousness effect* (as discussed in the preceding paragraph). Using ideas analogous to mediation analysis, they showed how the indirect effect can be decomposed into these two components in households of size two, assuming partial interference. Ogburn and Vander Weele [47] extended these methods to groups of arbitrary size and one large network without assuming partial interference.

### Other Approaches to Causal Inference with Interference

Most of the approaches described above utilize the potential outcomes framework for causal inference to assess treatment or exposure effects in the presence of interference. Randomization-based inference or large sample frequentist methods are typically employed for drawing statistical inferences. Below brief descriptions are provided of some recently proposed alternative approaches to causal inference in the presence of interference.

Carnegie et al. [55] considered an epidemic (mathematical) modeling approach for estimating the overall treatment effect in the presence of interference both within and across clusters in cluster-randomized trials of infectious disease.

Ogburn and Vander Weele [56] extended causal diagrams (or graphs) to the interference setting. They considered how causal graphs can help distinguish among causal mechanisms that give rise to interference, including the infectiousness and contagion effects described above.

Manski [57] considered decisions faced by a health planner concerned with vaccination policy, in particular whether to mandate vaccination or allow individuals to self-select whether to be vaccinated. An econometric perspective was adopted which includes consideration of the indirect effect of vaccination on persons not vaccinated. The health policy decision was framed as an optimization problem where a certain cost function is to be minimized which accounts for both the social cost of illness and the social cost of vaccination. Manski then considered the

optimal decision in settings where the indirect effect is assumed monotone with increasing vaccine coverage but otherwise unknown. Seasonal influenza vaccination was used as an example. This econometric perspective differs from the potential outcomes based approaches described above in that Manski did not seek to estimate or draw inference about the indirect effects of vaccination. Instead, the goal was to determine the optimal vaccination policy given only partial knowledge of the indirect effect.

Laber et al. [58] developed a Bayesian online estimation method to determine optimal treatment allocation strategies for control of white-nose syndrome in bats, a usually fatal disease caused by a fungus. The possible interference structure is described by a network where the nodes are counties in the USA. Interference and spillover effects are governed by a model describing the probability of spatial spread of white-nose syndrome between counties.

### Conclusions

One hundred years after publication of Ross' paper on dependent happenings, research on estimating treatment effects in the presence of interference is burgeoning. One current trend is to develop methods in the presence of general interference. These methods often involve using concepts from network theory. Applications motivating this work include social media, economics, and infectious diseases.

Future studies to evaluate causal effects under interference may continue to rely on cluster-randomization with the accompanying assumption of partial interference. But increasingly information on how people travel between clusters will be collected and analyzed to better understand contamination across clusters. Methods are needed that incorporate such information into both the causal estimands of interest and methods for inference. In the infectious disease context, genomic data is increasingly being gathered on the pathogen sequences. Phylogenetic information could also contribute to understanding the network structure of contacts and transmission. Much research remains to be done in developing methods for causal inference in the presence of interference.

**Acknowledgments** This work was partially funded by National Institutes of Health grants R01 AI085073 and R37 AI032042.

### Compliance with Ethical Standards

**Conflict of Interest** M. Elizabeth Halloran and Michael G. Hudgens declare a grant support from NIAID.

**Human and Animal Rights and Informed Consent** This article does not contain any studies with human or animal subjects performed by any of the authors.

## References

Papers of particular interest, published recently, have been highlighted as:

- Of importance

- Ross R. An application of the theory of probabilities to the study of *a priori* pathometry, part 1. *Proc R Soc Ser A*. 1916;92:204–30.
- Ross R. The prevention of malaria. 2nd ed. London: John Murray; 1911.
- Ross R. Some *a priori* pathometric equations. *Brit Med J*. 1915;1: 546–7.
- Holland PW. Statistics and causal inference. *J Am Stat Assoc*. 1986;81:945–60.
- Neyman JS. On the application of probability theory to agricultural experiments. Essay on principles. Section 9, translated by D.M. Dabrowska and T.P. Speed. *Stat Sci*. 1923; [reprinted 1990]5: 465–72.
- Robins JM. A new approach to causal inference in mortality studies with sustained exposure period—application to control of the healthy worker survivor effect. *Math Model*. 1986;7:1393–512.
- Rubin DB. Discussion of “Randomization analysis of experimental data in the Fisher randomization test” by Basu. *J Am Stat Assoc*. 1980;75:591–3.
- Cox DR. Planning of experiments. New York: John Wiley and Sons, Inc; 1958.
- Rubin DB. Bayesian inference for causal effects: the role of randomization. *Ann Stat*. 1978;7:34–58.
- Manski CF. Identification of treatment response with social interactions. *Econ J*. 2013;16:S1–S23.
- Halloran ME, Struchiner CJ. Study designs for dependent happenings. *Epidemiology*. 1991;2:331–8.
- Halloran ME, Struchiner CJ. Causal inference for infectious diseases. *Epidemiology*. 1995;6:142–51.
- Hudgens MG, Halloran ME. Towards causal inference with interference. *J Am Stat Assoc*. 2008;103:832–42.
- VanderWeele TJ, Tchetgen Tchetgen EJ. Effect partitioning under interference in two-stage randomized vaccine trials. *Stat Probab Lett*. 2011;81:861–9.
- P Toulis and E Kao. Estimation of causal peer influence effects. Proceedings of the 30th International Conference on Machine Learning. *JMLR W&CP*. 2013; 28.
- VanderWeele TJ, Tchetgen Tchetgen E, Halloran ME. Components of the indirect effect in vaccine trials: identification of contagion and infectiousness effects. *Epidemiology*. 2012;23(5):751–61.
- Rubin DB. Comment: Neyman (1923) and causal inference in experiments and observational studies. *Stat Sci*. 1990;5:472–80.
- Sobel M. What do randomized studies of housing mobility demonstrate? Causal inference in the face of interference. *J Am Stat Assoc*. 2006;101:1398–407.
- Tchetgen Tchetgen EJ, VanderWeele TJ. On causal inference in the presence of interference. *Stat Methods Med Res*. 2012;21(1):55–75. **Reviewed literature on causal inference in the presence of interference. Also developed new methods, including an inverse probability weighted estimator for observational data with partial interference.**
- Liu L, Hudgens MG. Large sample randomization inference of causal effects in the presence of interference. *J Am Stat Assoc*. 2014;109:288–301.
- Rigdon J, Hudgens MG. Exact confidence intervals in the presence of interference. *Stat Probab Lett*. 2015;105:130–5.
- S Baird, A Bohren, C McIntosh, and B Ozler. Designing experiments to measure spillover effects. PIER Working Paper 2014; 14–032. **Considered design of randomized experiments to draw inference about causal indirect (spillover), total, and overall effects in the context of economic studies with an empirical example.**
- Sur D, Ochiai R, Bhattacharya SK, Ganguly NK, Ali M, Manna B, Dutta S, Donner A, Kanungo S, Park JK, et al. A cluster-randomized effectiveness trial of Vi typhoid vaccine in India. *N Engl J Med*. 2009;361:335–44.
- Moulton LH, O’Brien KL, Kohberger R, Chang I, Reid R, Weatherholtz R, Hackell JG, Siber GR, Santosham M. Design of a group-randomised *Streptococcus pneumoniae* vaccine trial. *Control Clin Trials*. 2001;22:438–52.
- Hayes RJ, Moulton LH. Cluster randomised trials: a practical approach. Boca Raton, FL: Chapman and Hall/CRC; 2009.
- Halloran ME. The minicommunity for assessing indirect effects of vaccination. *Epidemiol Methods*. 2012;1(1):83–105.
- Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies. *Biometrika*. 1983;70:41–55.
- Hong S, Raudenbush G. Evaluating kindergarten retention policy: a case study of causal inference for multilevel observational data. *J Am Stat Assoc*. 2006;101:901–10.
- Lundin M, Karlsson M. Estimation of causal effects in observational studies with interference between units. *Statistical Methods & Applications* 2014;23:417–33.
- Perez-Heydrich C, Hudgens MG, Halloran ME, Clemens JD, Ali M, Emch ME. Assessing effects of cholera vaccination in the presence of interference. *Biometrics*. 2014;70(3):731–41.
- L Liu, MG Hudgens, and S Becker-Dreps. On inverse probability weighted estimators in the presence of interference. Unpublished manuscript, 2016.
- B Saul. interference: methods for causal inference with interference. R package version 0.4.62, <http://CRAN.R-project.org/package=inference>. 2015.
- Lazer D, Pentland A, Adamic L, et al. Computational social science. *Science*. 2009;323:721–3.
- Salathé M, Kazandjieva M, Lee JW, Levis P, Feldman MW, Jones JH. A high-resolution human contact network for infectious disease transmission. *Proc Natl Acad Sci*. 2010;107(51):22020–5.
- Wesolowski A, Metcalf CJE, Eagle N, et al. Quantifying seasonal population fluxes driving rubella transmission dynamics using mobile phone data. *Proc Natl Acad Sci*. 2015;112(35):11114–9.
- Balcan D, Colizza V, Gonçalves B, Hud H, Ramasco JJ, Vespignani A. Multiscale mobility networks and the spatial spreading of infectious diseases. *Proc Natl Acad Sci*. 2009;106(51):21484–9.
- Rosenbaum PR. Interference between units in randomized experiments. *J Am Stat Assoc*. 2007;1002:191–200.
- J Ugander, B Karrer, L Backstrom, and J Kleinberg. Graph cluster randomization: network exposure to multiple universes. 2013; arXiv:1305.6979. **Considered design of randomized experiments on one large network.**
- D Eckles, B Karrer, and J Ugander. Design and analysis of experiments in networks: reducing bias from interference. arXiv: 1404.7530, 2014.
- van der Laan MJ. Causal inference for a population of causally connected units. *J Causal Infer*. 2014;2:13–74. **Developed targeted maximum likelihood estimation for causal effects in the presence of general interference.**
- O Sofrygin and M van der Laan. Semi-parametric estimation and inference for the mean outcome of the single time-point intervention in a causally connected population. U.C. Berkeley Division of Biostatistics Working Paper Series. Working Paper 344. <http://biostats.bepress.com/ucbbiostat/paper344>, 2015.
- PM Aronow and C Samii. Estimating average causal effects under interference between units. 2015; arXiv:1305.6156v2:1–35.



- Developed a general approach for estimating effects of interventions in the presence of general forms of interference.**
43. Manski CF. Identification of endogenous social effects: the reflection problem. *Rev Econ Stud.* 1993;60(3):531–42.
  44. W An. Models and methods to identify peer effects. In: *the sage handbook of social network analysis.* London: Sage; 2011. pp. 515–532.
  45. VanderWeele TJ, An W. Social networks and causal inference. In: *handbook of causal analysis for social research.* Netherlands: Springer; 2013. p. 353–74.
  46. AJ O'Malley and JP Onnela. *Topics in social network analysis and network science.* 2014; arXiv:1404.0067.
  47. EL Ogburn and TJ VanderWeele. Vaccines, contagion, and social networks. 2014; arXiv:1403.1241.
  48. Préziosi M-P, Halloran ME. Effects of pertussis vaccination on transmission: vaccine efficacy for infectiousness. *Vaccine.* 2003;21:1853–61.
  49. Hudgens MG, Halloran ME. Causal vaccine effects on binary post-infection outcomes. *J Am Stat Assoc.* 2006;101:51–64.
  50. Frangakis CE, Rubin DB. Principal stratification in causal inference. *Biometrics.* 2002;58:21–9.
  51. VanderWeele TJ, Tchetgen Tchetgen EJ. Bounding the infectiousness effects in vaccine trials. *Epidemiology.* 2011;22(5):686–93.
  52. Halloran ME, Hudgens MG. Causal vaccine effects for infectiousness. *Int J Biostat.* 2012;8(2):Article 6.
  53. Halloran ME, Hudgens MG. Comparing bounds on vaccine effects on infectiousness. *Epidemiology.* 2012;23:931–2.
  54. VanderWeele TJ, Tchetgen Tchetgen E, Halloran ME. Interference and sensitivity analysis. *Stat Sci.* 2014;29(4):687–706.
  55. NB Carnegie, R Wang, and V De Gruttola. Estimation of the overall treatment effect in the presence of interference in cluster-randomized trials of infectious disease prevention. *Epidemiol Methods.* 2016, in press.
  - 56\*. Ogburn EL, VanderWeele TJ. Causal diagrams for interference. *Stat Sci.* 2014;29(4):559–78. **Extended causal graphs to allow for interference.**
  57. CF Manski. Mandating vaccination with unknown indirect effects. Working paper. 2016.
  58. EB Laber, NJ Meyer, BJ Reich, K Pacifici, JM Collazo, and JM Drake. On-line estimation of an optimal treatment allocation strategy for the control of white-nose syndrome in bats. Unpublished manuscript. 2016.