Randomization for the susceptibility effect of an infectious disease intervention

Daniel J. Eck1 · Olga Morozova2 · Forrest W. Crawford3,4,5,6

Received: 14 October 2020 / Revised: 7 June 2022 / Accepted: 5 July 2022 / Published online: 20 September 2022

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract

Randomized trials of infectious disease interventions, such as vaccines, often focus on groups of connected or potentially interacting individuals. When the pathogen of interest is transmissible between study subjects, interference may occur: individual infection outcomes may depend on treatments received by others. Epidemiologists have defined the primary parameter of interest—called the “susceptibility effect”—as a contrast in infection risk under treatment versus no treatment, while holding exposure to infectiousness constant. A related quantity—the “direct effect”—is defined as an unconditional contrast between the infection risk under treatment versus no treatment. The purpose of this paper is to show that under a widely recommended randomization design, the direct effect may fail to recover the sign of the true susceptibility effect of the intervention in a randomized trial when outcomes are contagious. The analytical approach uses structural features of infectious disease transmission to define the susceptibility effect. A new probabilistic coupling argument reveals stochastic dominance relations between potential infection outcomes under different treatment allocations. The results suggest that estimating the direct effect under randomization may provide misleading conclusions about the effect of an intervention—such as a vaccine—when outcomes are contagious. Investigators who estimate the direct effect may wrongly

1 Department of Statistics, University of Illinois Urbana-Champaign, Champaign, USA
2 Department of Public Health Sciences, Biological Sciences Division, The University of Chicago, Chicago, USA
3 Department of Biostatistics, Yale School of Public Health, New Haven, USA
4 Department of Statistics and Data Science, Yale University, New Haven, USA
5 Department of Ecology and Evolutionary Biology, Yale University, New Haven, USA
6 Yale School of Management, New Haven, USA
conclude an intervention that protects treated individuals from infection is harmful, or that a harmful treatment is beneficial.

**Keywords**  Contagion · Direct effect · Interference · Probabilistic coupling · Transmission model · Vaccine

**Mathematics Subject Classification**  92D30 · 92B05

## 1 Introduction

Randomized trials are widely used in the evaluation of infectious disease interventions among potentially interacting individuals (Halloran et al. 2010). For example, randomized trials have been employed to evaluate the effects of interventions, including vaccines, to prevent influenza (Belshe et al. 1998; Hayden et al. 2000; Welliver et al. 2001; Monto et al. 2002), pertussis (Simondon et al. 1997), and typhoid (Acosta et al. 2005), among many other diseases. The primary goal of most infectious disease intervention trials is to estimate the causal effect of treatment on the infection risk of the individual who receives it. However, when the infection is transmissible, or contagious, between study subjects, the treatment delivered to one subject may affect the infection outcome of others, via prevention of the original subject’s infection or reduction in their infectiousness once infected (Halloran and Struchiner 1995). A phenomenon known as “interference” arises when the treatment received by one individual exerts a causal effect on the outcomes of other individuals.

The “susceptibility effect” is of primary epidemiological interest in vaccine trials because it summarizes the effect of the intervention on the person who receives it, holding exposure to infection constant (O’Hagan et al. 2014). Halloran et al. (2010, page 19) write, “Historically, the primary focus has been how well vaccination protects the vaccinated individual. $V_{ES}$, the vaccine efficacy for susceptibility, is a measure of how protective vaccination is against infection”. The susceptibility effect is sometimes called the “susceptibility induced by the vaccine” (Becker and Britton 2004), the “conditional direct causal effect” (Halloran and Struchiner 1995), or per-exposure effect (O’Hagan et al. 2014), and may be represented by a hazard ratio, risk ratio, or risk difference (Cole et al. 2015). Unfortunately, the susceptibility effect can be difficult to estimate because exposure to infection cannot always be precisely measured.

A related quantity, called the “direct effect”, is defined as an unconditional contrast between infection outcomes among treated and untreated individuals (Halloran et al. 2010). Hudgens and Halloran (2008) proposed a randomization design and a definition of the “direct effect” under interference in a clustered study population, along with effect estimators. Informally, the direct effect is defined as a contrast between the rate of infection for an individual under treatment versus no treatment, averaged over the conditional distribution of treatments to others in the same cluster (VanderWeele and Tchetgen 2011; Sävje et al. 2021). The direct effect estimand introduced by Hudgens and Halloran (2008) has been applied in empirical analyses of randomized trials (e.g. Buchanan et al. 2018).
The susceptibility effect and direct effect are not the same. However, they may appear to measure similar causal features of the effect of an intervention on individuals who receive it, especially under randomization. Randomization ensures that on average, treated and untreated individuals do not vary systematically in their baseline characteristics. Indeed, Halloran and Struchiner (1995, page 146) write “Under a random assignment of the vaccine to the population, then if everyone were exposed to infection, the average causal direct effect of the vaccine on the transmission probability would be estimated as the difference in the average outcomes in the unvaccinated and vaccinated individuals under the actual treatment assignment”. But even when treatment is randomized, exposure to infection can be systematically different among treated and untreated individuals during the study. Researchers have warned that this differential exposure can confound estimates of the “direct effect” of the intervention (Halloran and Struchiner 1995; Halloran et al. 2010; Kenah 2014; Morozova et al. 2018; Cai et al. 2021), but the relationship between the randomization design and the disease transmission process remains obscure (Struchiner and Halloran 2007; van Boven et al. 2013; O’Hagan et al. 2014). Do contrasts of infection outcomes between treated and untreated subjects, as proposed by Hudgens and Halloran (2008) as the “direct effect”, recover the susceptibility effect of the intervention when the population is clustered, treatment is randomized, and outcomes are contagious?

The purpose of this paper is to examine the meaning of the “direct effect” defined by Hudgens and Halloran (2008) when infectious disease outcomes are transmissible in a study of potentially interacting individuals within a clustered study population (Becker et al. 2003, 2006). We first provide a formal definition of the causal susceptibility effect, which is of primary interest in trials of infectious disease interventions. We then briefly review the direct effect, and define three common randomization designs—Bernoulli, block, and cluster randomization—that may be employed in empirical trials of infectious disease interventions. To compare the susceptibility and direct effects in a trial of an infectious disease intervention, we evaluate infection outcomes under a general structural model of infectious disease transmission in clusters that accommodates individually varying susceptibility to infection, infectiousness, and exogenous source of infection. The structural model is based on the canonical stochastic susceptible-infective epidemic process (Becker 1989; Andersson and Britton 2000; Diekmann et al. 2012). This type of structural model has found wide application in studies of infectious disease outcomes in clusters of individuals (Rhodes et al. 1996; Longini Jr et al. 1999; Auranen et al. 2000; O’Neill et al. 2000; Becker et al. 2003; Becker and Britton 2004; Cauchemez et al. 2004, 2006, 2009; Becker et al. 2006; Kenah 2013, 2014; Tsang et al. 2015, 2016; Morozova et al. 2018). We show that under some forms of randomization, the direct effect may not recover the sign of the true susceptibility effect of the intervention on the individual who receives it. In particular, when the intervention both helps protect treated individuals from infection, and helps prevent infected treated individuals from transmitting the infection to others, the direct effect can nevertheless be positive (indicating harm) under the randomization design proposed by Hudgens and Halloran (2008). The results are derived using a probabilistic coupling (den Hollander 2012; Ross 1996) argument that reveals stochastic dominance relations between infection outcomes under different treatment allocations. These results substantially sharpen the claims of Halloran et al. (1991) and
those made by VanderWeele and Tchetgen (2011) and Sävje et al. (2021) who also point out difficulties with the interpretation of the direct effect defined by Hudgens and Halloran (2008) in the presence of interference, and generalize bias results for clusters of size two (Morozova et al. 2018).

2 Setting

Consider a population of $N$ clusters, and let $n_i$ be the number of individuals in cluster $i$. Suppose the outcome of interest is infection by an infectious disease that is transmissible between individuals within clusters, but not between clusters. Let $T_{ij}$ be the random infection time of subject $j$ and let $Y_{ij}(t) = 1 \{ T_{ij} < t \}$ be the indicator of prior infection. A subject $j$ is called susceptible at time $t$ if $Y_{ij}(t) = 0$ and infected if $Y_{ij}(t) = 1$. The joint treatment vector $x_i = (x_{i1}, \ldots, x_{in_i})$ is allocated at baseline, $t = 0$. We will sometimes write the joint treatment allocation in cluster $i$ as $x_i = (x_{ij}, x_{i(j)})$, where $x_{ij}$ is the treatment to subject $j$, and $x_{i(j)}$ is the vector of treatment assignments to subjects other than $j$ in cluster $i$.

2.1 Target parameter: susceptibility effect

To define meaningful intervention effects for infectious disease outcomes, it is often necessary to consider a joint intervention on both the treatment assignment and exposure history of cluster members (O’Hagan et al. 2014). We use potential outcome notation (Rubin 2005) to define causal effects. Let the infection status history of all subjects other than $j$ in cluster $i$ be denoted $H_{i(j)} = \{ Y_{ik}(s), \ k \neq j, \ s \geq 0 \}$, with a particular realization denoted by $h_{i(j)}$. The infection history $H_{i(j)}$ is a vector of $n_i - 1$ indicator functions denoting infection status for all times $0 \leq s < \infty$. Let $T_{ij}(x_i, h_{i(j)})$ be the potential infection time of $j$ when treatment is set to $x_i$ and the infection history of individuals other than $j$ is set to $h_{i(j)}$. Let $Y_{ij}(t, x_i, h_{i(j)}) = 1 \{ T_{ij}(x_i, h_{i(j)}) < t \}$ be the corresponding potential infection outcome of subject $j$ at time $t$. It is implicit that for fixed $t$, the potential infection outcome $Y_{ij}(t, x_i, h_{i(j)})$ does not depend on any element $k$ of $h_{i(j)}$ when $T_{ik} > t$. In other words, infection of $k$ after $t$ does not affect infection of $j$ prior to $t$. When $x_i$ is a fixed treatment allocation and the infection history of other individuals $H_{i(j)}$ is allowed to arise naturally without intervention on infection history, we write $T_{ij}(x_i) = T_{ij}(x_i, H_{i(j)})$ and $Y_{ij}(t, x_i) = 1 \{ T_{ij}(x_i, H_{i(j)}) < t \}$. We regard potential infection outcomes as inherently stochastic (VanderWeele and Robins 2012): given a treatment allocation $x_i$ and infection histories $h_{i(j)}$, the potential infection time $T_{ij}(x_i, h_{i(j)})$ is a random variable.

Following O’Hagan et al. (2014), we define the susceptibility effect as a contrast of the infection outcome of $j$ under treatment ($x_{ij} = 1$) versus no treatment ($x_{ij} = 0$), while holding constant the treatments $x_{i(j)}$ and infection histories $H_{i(j)}$ of other cluster members. Define the potential hazard of infection to subject $j$ in cluster $i$ at time $t$ as the instantaneous risk of infection at time $t$, given no infection up to $t$, holding other individuals’ infection history $h_{i(j)}$ and treatments $x_{i(j)}$ constant:

$$\lambda_{ij}(t, x_{ij}, x_{i(j)}, h_{i(j)}) = \lim_{\epsilon \to 0^+} \frac{\mathbb{E}\left[ Y_{ij}(t + \epsilon, x_{ij}, x_{i(j)}, h_{i(j)}) | Y_{ij}(t, x_{ij}, x_{i(j)}, h_{i(j)}) = 0 \right]}{\epsilon}$$
when this limit exists. The susceptibility hazard ratio (SHR) contrasts potential hazards under treatment versus no treatment of \( j \), while holding all else—including exposure to infection—constant (Halloran et al. 1997):

\[
\text{SHR}_{ij} (t, x_{i(j)}, h_{i(j)}) = \frac{\lambda_{ij} (t, 1, x_{i(j)}, h_{i(j)})}{\lambda_{ij} (t, 0, x_{i(j)}, h_{i(j)})}. \tag{1}
\]

Informally, \( \text{SHR}_{ij} \) contrasts the instantaneous potential risk of infection of susceptible subject \( j \) at time \( t \) under treatment versus no treatment, while holding constant exposure to infection determined by the treatments and infection histories of others. Halloran et al. (1997, Table 1) define the “vaccine effect on susceptibility” as \( 1 - \text{SHR}_{ij} \). Cluster and population-level susceptibility estimands may be defined as expectation of \( \text{SHR}_{ij} (t, X_{i(j)}, H_{i(j)}) \), or as a ratio of expectations of the hazards. Analogous exposure-conditioned susceptibility effects can be defined on the risk difference and odds ratio scales (O’Hagan et al. 2014).

2.2 The “direct effect” in a randomized trial

Define the expected individual infection outcome under joint treatment \( x_i \) as \( \bar{Y}_{ij} (t, x_i) = \mathbb{E}[Y_{ij} (t, x_i)] \), where expectation is with respect to the infection outcomes in cluster \( i \). Let \( \mathcal{X}^n = \{0, 1\}^n \) be the set of all binary vectors of \( n \) elements. We define causal estimands by comparing average infection outcomes under different treatment allocations to the cluster. These definitions are taken, with minor changes in notation, from Hudgens and Halloran (2008). Define the individual average potential outcome as

\[
\bar{Y}_{ij} (t, x, x_{i(j)}) = \sum_{x_{i(j)} \in \mathcal{X}^n} \bar{Y}_{ij} (t, x, x_{i(j)}) \Pr(X_{i(j)} = x_{i(j)} | X_{ij} = x). \tag{2}
\]

Informally, \( \bar{Y}_{ij} (t, x, x_{i(j)}) \) is the individual infection outcome under \( x_{ij} = x \), averaged over the conditional distribution of treatments to the other individuals in cluster \( i \). Define the cluster average potential outcome as \( \bar{Y}_i (t, x) = n_i^{-1} \sum_{j=1}^{n_i} \bar{Y}_{ij} (t, x) \), and the population average potential outcome as \( \bar{Y} (t, x) = N^{-1} \sum_{i=1}^{N} \bar{Y}_i (t, x) \). Hudgens and Halloran (2008) propose contrasts of these potential outcomes as causal estimands, which we rewrite in slightly different form. Define the individual average direct effect as \( DE_{ij} (t) = \bar{Y}_{ij} (t, 1) - \bar{Y}_{ij} (t, 0) \), the cluster average direct effect as \( DE_i (t) = n_i^{-1} \sum_{j=1}^{n_i} DE_{ij} (t) \), and the population average direct effect as \( DE (t) = N^{-1} \sum_{i=1}^{N} DE_i (t) \). VanderWeele and Tchetgen (2011) and Sävje et al. (2021) point out issues with the interpretation of the direct effect defined by Hudgens and Halloran (2008), and propose alternative definitions and randomization schemes.

2.3 Randomization designs for clustered subjects

A randomization design is a probability distribution that assigns the joint binary treatment vector \( x_i = (x_{i1}, \ldots, x_{in_i}) \) within and across clusters.
Definition 1 (Bernoulli randomization) The treatment is Bernoulli randomized if for every cluster $i$, the joint allocation $x_i = (x_{i1}, \ldots, x_{in_i})$ has probability $\Pr(X_i = x_i) = \prod_{j=1}^{n_i} p^{x_{ij}}(1 - p)^{1-x_{ij}}$ for some probability $p$.

Definition 2 (Block randomization) The treatment is block-randomized if for every cluster $i$, the joint allocation $x_i = (x_{i1}, \ldots, x_{in_i})$ has probability $\Pr(X_i = x_i) = \frac{\binom{n_i}{m_i}}{\binom{n_i}{\lfloor pn_i \rfloor}}$ where $0 < m_i = \lfloor pn_i \rfloor$ for some probability $p > \frac{1}{\min_i n_i}$.

Definition 3 (Cluster randomization) The treatment is cluster randomized if for each cluster $i$, either all members of the cluster are treated, or all are untreated with probability $0 < p < 1$. That is, $\Pr(X_i = (1, \ldots, 1)) = p$ and $\Pr(X_i = (0, \ldots, 0)) = 1 - p$ for each cluster $i$ independently.

Block and cluster randomization designs induce dependencies in the treatment status of subjects in the same cluster. This means that the conditional treatment probability $\Pr(X_{i(j)} = x_{i(j)} | X_{ij} = x)$ in (2) may differ for $x = 1$ and $x = 0$, and so the individual average risk difference $DE_{ij}(t)$ may not be an average of individual effects. VanderWeele and Tchetgen (2011) point out that the risk difference $DE(t)$ may suffer from difficulties in interpretation under block randomization. Sävje et al. (2021) call $DE(t)$ the “average distribution shift effect” because it “captures the compound effect of changing a unit’s treatment and simultaneously changing the experimental design”. However, it remains unclear whether the direct effect $DE_{ij}(t)$ has a meaningful interpretation when interference arises via contagion.

3 Approach

Do the “direct effect” quantities $DE_{ij}(t)$, $DE_i(t)$, and $DE(t)$ above recover useful features of the susceptibility effect of interest in a randomized trial? For example, if the treatment $x$ is a vaccine that truly helps prevent infection in the person who receives it when exposure to infection is held constant ($SHR < 1$), investigators conducting a randomized trial might want to know whether they should expect $DE(t) < 0$. To answer this question, we must specify more precisely the way that infection outcomes arise under contagion. Epidemiologists have proposed structural models of infectious disease outcomes that formalize common ideas about the mechanism, or dynamics, of transmission in groups (Becker 1989; Anderson and May 1992; Andersson and Britton 2000). Many structural transmission models represent the individual risk (or hazard) of infection as an explicit function of individual treatments and possibly other covariates (Auranen et al. 2000; O’Neill et al. 2000; Becker et al. 2003; Becker and Britton 2004; Becker et al. 2006; Cauchemez et al. 2004, 2006; Kenah 2013, 2014; Morozova et al. 2018). Structural models can be useful in both observational and randomized trials because they posit an explicit regression-style relationship linking covariates and infection outcome.

We present a general structural model of infectious disease transmission based on the canonical stochastic susceptible-infective epidemic process (Becker 1989; Andersson and Britton 2000; Diekmann et al. 2012). This model, based on constructions
by Kenah (2014), captures the essential features of infectious disease transmission, and the effect of treatment on susceptibility and infectiousness. In particular, this model represents the instantaneous risk (hazard) of infection experienced by subject $j$ in cluster $i$ as a non-decreasing step function in time $t$ whose jumps correspond to infections of other cluster members. Conveniently, the susceptibility effect $\text{SHR}$ corresponds explicitly to a parameter in this model. Recall that $h_{i(j)}$ consists of the infection histories of individuals other than $j$: $h_{i(j)} = \{y_{ik}(t), k \neq j, t \geq 0\}$. Let the hazard of infection experienced by a susceptible individual $j$ in cluster $i$ at time $t$ be

$$
\lambda_{ij}(t, x_i, h_{i(j)}) = e^{x_{ij}\beta + \eta_{ij}} \left( \alpha + \sum_{k=1}^{n_i} y_{ik}(t)e^{x_{ik}\gamma + \xi_{ik}} \right)
$$

where $\beta$ is the effect of individual treatment $x_{ij}$, $\eta_{ij}$ is an individual susceptibility coefficient, $\alpha$ is the force of infection from outside the cluster, $\gamma$ is the infectiousness effect of the treatment $x_{ik}$ assigned to $k$ and $\xi_{ik}$ is an individual infectiousness coefficient for $k$. The sum over $k$ in (3) does not include $k = j$ because $j$ cannot infect themselves. Under this structural model, the susceptibility effect of interest (1) has a simple time-invariant form: $\text{SHR}_{ij}(t, x_i(j), h_{i(j)}) = e^\beta$.

The structural transmission model (3) formalizes intuition about how interference arises for infectious disease outcomes. The hazard of infection experienced by subject $j$ at time $t$ is a function of subject $k$’s features ($x_{ik}$ and $\xi_{ik}$) only when $k$ is currently infected ($y_{ik}(t) = 1$). The structural transmission model (3) obeys “partial interference” (Sobel 2006): the infection outcome for subject $j$ in cluster $i$ may depend on treatments and infection outcomes of other individuals in cluster $i$, but does not depend on subjects in clusters other than $i$. Variations on this infection hazard model (3) have been used to model sources of disease transmission and for estimation of covariate effects on infection risk (Auranen et al. 2000; Cauchemez et al. 2004, 2006; Kenah 2013, 2014), and as a conceptual model to evaluate the properties of risk ratios under contagion (Morozova et al. 2018). Figure 1 shows a schematic illustration of the transmission hazard model (3) for a cluster $i$ of size $n_i = 4$ in which two subjects are treated.

4 Results

4.1 DE under the null hypothesis of no susceptibility effect

If the direct effect is to serve as a useful estimand for researchers interested in learning about the causal effect of the intervention on the subject who receives it, we should expect that $DE_{ij}(t) = 0$ when $\beta = 0$, since the treatment has no effect on the infection risk of an individual who receives it. We begin by studying the properties of the average individual direct effect $DE_{ij}(T)$ under the three randomization designs. We assume that the exogenous (community) force of infection $\alpha$ is positive, and $T > 0$ is a follow-
up time at which infection outcomes are measured, so that at least one infection in each cluster arises with positive probability.

Bernoulli randomization gives concordance between $\beta = 0$ and the direct effect.

**Proposition 1 (DE under Bernoulli randomization)** Suppose $\beta = 0$ and treatment assignment is Bernoulli randomized. Then $DE_{ij}(T) = 0$.

In contrast, the direct effect has the opposite sign as the infectiousness effect $\gamma$ when $\beta = 0$ under block randomization.

**Proposition 2 (DE under block randomization)** Suppose $\beta = 0$ and treatment assignment is block-randomized. If $\gamma < 0$ then $DE_{ij}(T) > 0$; if $\gamma = 0$ then $DE_{ij}(T) = 0$; and if $\gamma > 0$ then $DE_{ij}(T) < 0$.

The direct effect has the same sign as $\gamma$ when $\beta = 0$ under cluster randomization.

**Proposition 3 (DE under cluster randomization)** Suppose $\beta = 0$ and treatment assignment is cluster randomized. If $\gamma < 0$ then $DE_{ij}(T) < 0$; if $\gamma = 0$ then $DE_{ij}(T) = 0$; and if $\gamma > 0$ then $DE_{ij}(T) > 0$.

Propositions 1–3 compare averaged expectations of infection outcomes for subject $j$ in cluster $i$. However, computing the expectation $\bar{Y}_{ij}(t, x, x_{i(j)})$ for particular values of $x$ and $x_{i(j)}$ is intractable, so an explicit comparison of average individual potential infection outcomes under different treatment allocations cannot be made analytically. Instead, we will use tools from the theory of probabilistic coupling (den Hollander 2012; Ross 1996) to exhibit stochastic dominance relations between infection outcomes under different treatment allocations to facilitate the comparison.
Definition 4 (Coupling) A coupling of two random variables \( Y^0 \) and \( Y^1 \) both taking values in \((\Omega, \mathcal{F})\) is any pair of random variables \((\tilde{Y}^0, \tilde{Y}^1)\) taking values in \((\Omega \times \Omega, \mathcal{F} \otimes \mathcal{F})\) whose marginal distributions are identical to those of \( Y^0 \) and \( Y^1 \) respectively, i.e. \( Y^0 \overset{d}{=} \tilde{Y}^0 \) and \( Y^1 \overset{d}{=} \tilde{Y}^1 \).

Typically the variables \( \tilde{Y}^0 \) and \( \tilde{Y}^1 \) are dependent. To study the relationship of infection outcomes under different treatment scenarios, a notion of dominance will be necessary.

Definition 5 (Stochastic dominance) The real-valued random variable \( Y^1 \) stochastically dominates \( Y^0 \) if \( \Pr(Y^1 < y) \leq \Pr(Y^0 < y) \) for all \( y \in \mathbb{R} \).

If \( Y^1 \) stochastically dominates \( Y^0 \) and vice versa, the variables are equal in distribution. If \( Y^1 \) stochastically dominates \( Y^0 \), then \( \mathbb{E}[Y^1] \geq \mathbb{E}[Y^0] \). The following Lemma, proved by e.g. Ross (1996, pages 409–410), provides a framework for establishing stochastic dominance through the construction of a coupling.

Lemma 1 (Coupling and stochastic dominance) The real-valued random variable \( Y^1 \) stochastically dominates \( Y^0 \) if and only if there is a coupling \((\tilde{Y}^0, \tilde{Y}^1)\) of \( Y^0 \) and \( Y^1 \) such that \( \Pr(\tilde{Y}^1 \geq \tilde{Y}^0) = 1 \).

To begin proving Propositions 1–3, define the vectors of stochastic potential outcomes of all subjects under two different joint treatments allocations \( x^1 \) and \( x^0 \) as \( Y_i(t, x^1_i) = (Y_{i1}(t, x^1_i), \ldots, Y_{ini}(t, x^1_i)) \) and \( Y_i(t, x^0_i) = (Y_{i1}(t, x^0_i), \ldots, Y_{ini}(t, x^0_i)) \). Corresponding to these potential outcomes, we will construct two coupled outcome processes with \( \beta = 0 \), denoted \( \tilde{Y}^1_i(t) \) and \( \tilde{Y}^0_i(t) \), under treatment vectors \( x^1_i \) and \( x^0_i \) respectively. The order of infections in both processes is the same, but the times of infection may be different.

Let \( S_l \) and \( I_l \) be the set of subjects that are susceptible and infectious, respectively, just before the \( l \)th infection. Let \( \tilde{W}^1_l \) and \( \tilde{W}^0_l \) be the waiting times to the next infection in the coupled processes under \( x^1_i \) and \( x^0_i \) respectively. Define the waiting time cumulative distribution functions

\[
F_l(w) = 1 - \exp \left[ -w \sum_{a \in S_l} e^{\eta_{ia}} \left( \alpha + \sum_{b \in I_l} e^{\gamma x^1_{ib} + \xi_{ib}} \right) \right]
\]

and

\[
G_l(w) = 1 - \exp \left[ -w \sum_{a \in S_l} e^{\eta_{ia}} \left( \alpha + \sum_{b \in I_l} e^{\gamma x^0_{ib} + \xi_{ib}} \right) \right],
\]

where sums over empty sets are interpreted as zero. Let \( \tilde{t}^1_{il} \) and \( \tilde{t}^0_{il} \) be the time of infection of subject \( l \) under treatments \( x^1_i \) and \( x^0_i \) respectively, with \( \tilde{t}^1_{il} = \tilde{t}^0_{il} = 0 \). Likewise define the corresponding infection indicators \( \tilde{Y}^1_{il}(t) = 1 \{ \tilde{t}^1_{il} < t \} \) and \( \tilde{Y}^0_{il}(t) = 1 \{ \tilde{t}^0_{il} < t \} \). The following algorithm constructs the joint outcome functions.
$\bar{Y}_i^1(t) = (\bar{Y}_i^1(t), \ldots, \bar{Y}_{in_i}^1(t))$ and $\bar{Y}_i^0(t) = (\bar{Y}_{i1}^0(t), \ldots, \bar{Y}_{in_i}^0(t))$ under treatment vectors $x_i^1$ and $x_i^0$ respectively. We show below that $(\bar{Y}_i^1(t), \bar{Y}_i^0(t))$ is a coupling of the potential infection outcomes $Y_i(t, x^1)$ and $Y_i(t, x^0)$.

**Algorithm 1** Construction of the coupling for cluster $i$.

```
S_l ← \{1, \ldots, n_i\}  \quad \triangleright \text{Initialize susceptibles}
I_l ← \emptyset  \quad \triangleright \text{Initialize infectives}

for $l ← 1, \ldots, n_i$ do
    $U_l \sim \text{Uniform}(0, 1)$
    $\bar{W}_l^1 ← F_i^{-1}(U_l)$
    $\bar{W}_l^0 ← G_i^{-1}(U_l)$
    for $v \in S_l$ do
        $p_v ← e^{\eta_v} / \sum_{a \in S_l} e^{\eta_a}$
    end for
    $V_l \sim \text{Multinomial}(S_l, \{p_v : v \in S_l\})$  \quad \triangleright \text{Choose next infected subject}
    $I_{l|V_l} ← I_{l|V_l-1} + \bar{W}_l^1$
    $I_{l|V_l} ← I_{l|V_l-1} + \bar{W}_l^0$
    $\tilde{Y}_i^1(t) ← 1\{\bar{I}_{l|V_l} < t\}$  \quad \triangleright \text{Define infection indicator function}
    $\tilde{Y}_i^0(t) ← 1\{\bar{I}_{l|V_l} < t\}$
    $S_l ← S_{l-1} \setminus \{V_l\}$
    $I_l ← I_{l-1} \cup \{V_l\}$
end for
```

Algorithm 1 generates two sets of infection outcomes, one corresponding to the joint treatment $x_i^1$ and one to the joint treatment $x_i^0$, by constructing waiting times to infection of each subject, and which subject is infected at each step. The key insight is that under the infection hazard model (3), the waiting times $\bar{W}_l^1$ and $\bar{W}_l^0$ depend on treatments of already-infected individuals, but because $\beta = 0$, selection of the next infected individual does not depend on treatments of yet-uninfected subjects. This fact permits construction of two dependent infection processes whose timing differs, but where the order of infections is identical.

**Lemma 2** (Construction of the coupling) When $\beta = 0$, the variables $(\bar{Y}_i^1(t), \bar{Y}_i^0(t))$ constructed by Algorithm 1 constitute a coupling of the potential infection outcomes $Y_i(t, x^1)$ and $Y_i(t, x^0)$.

**Proof of Lemma 2** We will show that $(\bar{Y}_i^1(t), \bar{Y}_i^0(t))$ is a coupling of $Y_i(t, x^1)$ and $Y_i(t, x^0)$ satisfying Definition 4. First, the waiting time distribution functions $F_i(w)$ and $G_i(w)$ are monotonically increasing in $w$, so the random waiting time $\bar{W}_l^1 = F_i^{-1}(U_l)$ has distribution function $F_i(w)$ and $\bar{W}_l^0 = G_i^{-1}(U_l)$ has distribution function $G_i(w)$ (Devroye 1986). Because the same uniform variable $U_l$ is used to generate both waiting times $\bar{W}_l^1$ and $\bar{W}_l^0$, these variables, and hence the infection times $\bar{I}_{l|V_l}^1$ and $\bar{I}_{l|V_l}^0$, and outcomes $\tilde{Y}_i^1(t)$ and $\tilde{Y}_i^0(t)$, are dependent. The joint mass function of the $l$th infected subject $V_l$ and the cumulative distribution function of the waiting time $\bar{W}_l^1$ to
this infection is, by construction,

\[ \Pr(V_l = v, \tilde{W}_l^1 < w) = \frac{e^{\eta_V v}}{\sum_{a \in S_l} e^{\eta_V a}} \left[ 1 - \exp \left[ -w \sum_{a \in S_l} e^{\eta_V a} \left( \alpha + \sum_{b \in I_l} e^{\gamma x_{ib} + \xi_{ib}} \right) \right] \right]. \]  

(4)

Differentiating (4) with respect to \( w \), we find that the joint likelihood of the newly infected subject \( V_l = v \) and the waiting time \( w \) to the \( l \)th infection is

\[ \tilde{L}_i^1(v, w) = e^{\eta_V v} \left( \alpha + \sum_{b \in I_l} \tilde{\gamma}_{ib} e^{\gamma x_{ib} + \xi_{ib}} \right) \exp \left[ -w \sum_{a \in S_l} e^{\eta_V a} \left( \alpha + \sum_{b \in I_l} e^{\gamma x_{ib} + \xi_{ib}} \right) \right] \]

\[ = e^{\eta_V v} \left( \alpha + \sum_{b = 1}^{n} \tilde{\gamma}_{ib} (\tilde{t}_{iv}) e^{\gamma x_{ib} + \xi_{ib}} \right) \exp \left[ -w \sum_{a = 1}^{n} (1 - \tilde{y}_{ia} (\tilde{t}_{iv})) e^{\eta_V a} \right] \]

\[ = \tilde{\lambda}_{iv}^1 (\tilde{t}_{iv}) \exp \left[ -w \sum_{a = 1}^{n} (1 - \tilde{y}_{ia} (\tilde{t}_{iv})) \tilde{\lambda}_{ia}^1 (\tilde{t}_{iv}) \right] \]  

(5)

where \( \tilde{\lambda}_{ij}^1 (t) \) is (3) with \( \tilde{y}_{ij}^1 (t) \) and \( x_i^1 \) replacing \( y_{ij} (t) \) and \( x_i \) respectively. Let \( \tilde{L}_i^1 (\tilde{y}_{ij}^1) \) be the likelihood of the full realization of \( \tilde{y}_{ij}^1 (t) = (\tilde{y}_{i1}^1 (t), \ldots, \tilde{y}_{in_i}^1 (t)) \) with \( \tilde{T}^1 = (\tilde{t}_{i1}^1, \ldots, \tilde{t}_{in_i}^1), \tilde{W}_i^1 = (\tilde{w}_{i1}^1, \ldots, \tilde{w}_{in_i}^1) \), and \( \tilde{V} = (\tilde{v}_{i1}^1, \ldots, \tilde{v}_{in_i}^1) \). Recall that by construction, \( \tilde{w}_{ik}^1 = \tilde{t}_{iv}^1 - \tilde{t}_{iv,k-1}^1 \). The likelihood of the constructed process is

\[ \tilde{L}(\tilde{y}_{ij}^1) = \prod_{k = 1}^{n_i} \tilde{\lambda}_{iv(k)}^1 (\tilde{t}_{iv(k)}) \exp \left[ -w_k \sum_{j = 1}^{n_i} (1 - \tilde{y}_{ij}^1 (\tilde{t}_{iv(k)}) \tilde{\lambda}_{ij}^1 (\tilde{t}_{iv(k)}) \right] \]

\[ = \left( \prod_{j = 1}^{n_i} \tilde{\lambda}_{ij}^1 (\tilde{t}_{ij}) \right) \exp \left[ -\sum_{j = 1}^{n_i} \sum_{k = 1}^{n_i} \int_{\tilde{t}_{iv(k-1)}}^{\tilde{t}_{iv(k)}} \left( 1 - \tilde{y}_{ij}^1 (t) \right) \tilde{\lambda}_{ij}^1 (t) \, dt \right] \]  

(6)

where \( L(\tilde{y}_{ij}^1 (t)) \) is the likelihood of the original process. Therefore the constructed outcome vector \( \tilde{Y}_i (t, x_i^1) \) is equal in distribution to the potential outcome vector \( Y_i (t, x_i^1) \), and it follows that \( \tilde{Y}_{ij}^1 (t, x_i^1) \) is equal in distribution to \( Y_{ij}^1 (t, x_i^1) \). By the same reasoning, \( \tilde{Y}_{ij}^1 (t, x_i^0) \) is equal in distribution to \( Y_{ij}^1 (t, x_i^0) \). Therefore by Definition 4, \( (\tilde{Y}_{ij}^1 (t, x_i^1), \tilde{Y}_{ij}^1 (t, x_i^0)) \) is a coupling of \( Y_{ij}^1 (t, x_i^1) \) and \( Y_{ij} (t, x_i^0) \).
Proof of Proposition 1 First, let \( \mathbf{x}_i^1 = (x_{ij} = 1, \mathbf{x}_{i(j)}) \) and \( \mathbf{x}_i^0 = (x_{ij} = 0, \mathbf{x}_{i(j)}) \) be joint treatment allocations that are identical except that for subject \( j \), \( x_{ij}^0 = 1 \) and \( x_{ij}^0 = 0 \). Then by Lemma 2, \( (\tilde{Y}_i^1(t), \tilde{Y}_i^0(t)) \) is a coupling of \( Y_i(t, \mathbf{x}_i^1) \) and \( Y_i(t, \mathbf{x}_i^0) \) under \( \beta = 0 \). Whenever \( j \) is uninfected, \( F_i(w) = G_i(w) \), so \( \tilde{T}_i^1 = \tilde{T}_i^0 \) and so \( \tilde{Y}_i^1(t) = \tilde{Y}_i^0(t) \). Therefore by Lemma 1 \( Y_{ij}(t, \mathbf{x}_i^1) \) stochastically dominates \( Y_{ij}(t, \mathbf{x}_i^0) \) and vice versa. It follows that \( Y_{ij}(t, \mathbf{x}_i^1) \) and \( Y_{ij}(t, \mathbf{x}_i^0) \) are equal in distribution, so \( \overline{Y}_{ij}(T, 1, \mathbf{x}_{i(j)}) = \overline{Y}_{ij}(T, 0, \mathbf{x}_{i(j)}) \). Now consider a Bernoulli randomized treatment allocation \( \mathbf{X}_{i(j)} \) to subjects other than \( j \) in cluster \( i \). Under Bernoulli randomization, the distribution of \( \mathbf{X}_{i(j)} \) is invariant to conditioning on \( X_{ij} = x_{ij} \). By the definition of the individual average potential infection outcome \( (2) \),

\[
\overline{Y}_{ij}(T, 1) = \sum_{\mathbf{x}_{i(j)} \in \mathcal{X}_{ni}^{n_i-1}} \overline{Y}_{ij}(T, 1, \mathbf{x}_{i(j)}) \prod_{k \neq j} p^{x_{ik}}(1 - p)^{1 - x_{ik}}
\]

\[
= \sum_{\mathbf{x}_{i(j)} \in \mathcal{X}_{ni}^{n_i-1}} \overline{Y}_{ij}(T, 0, \mathbf{x}_{i(j)}) \prod_{k \neq j} p^{x_{ik}}(1 - p)^{1 - x_{ik}}
\]

\[
= \overline{Y}_{ij}(T, 0),
\]

and so \( DE_{ij}(T) = 0 \) as claimed. \( \square \)

The proof of Proposition 2 proceeds similarly, but we evaluate differences in potential infection outcomes of subject \( j \) when \( j \) and \( k \neq j \) have opposite treatments, with other subjects’ treatments held constant.

Proof of Proposition 2 Let \( \mathcal{X}_m^n \) be the set of all binary \( n \)-vectors with \( m \) positive elements. First, we deduce a stochastic order relation for a particular treatment allocation in which \( j \) and \( k \) have opposite treatments. Let \( \mathbf{z} \in \mathcal{X}_{m_{j(k)}-1}^{n_{j(k)}} \) and for \( j \neq k \) define \( \mathbf{x}_i = (x_{ij} = 1, x_{ik} = 0, \mathbf{x}_{ij(k)} = \mathbf{z}) \) and \( \mathbf{x}_j = (x_{ij} = 0, x_{ik} = 1, \mathbf{x}_{ij(k)} = \mathbf{z}) \). When \( \gamma = 0 \), \( F_i(w) = G_i(w) \) for all \( l \) and all \( w \). Therefore \( \tilde{T}_{i(l)} = \tilde{T}_{i0} \) for all \( l \) and so \( \tilde{Y}_{ij}(t) = \tilde{Y}_{ij}(t) \) for all \( t \). Then \( Y_{ij}(t, \mathbf{x}_i^1) \) is equal in distribution to \( Y_{ij}(t, \mathbf{x}_i^0) \) for all \( t \) and so \( \overline{Y}_{ij}(t, \mathbf{x}_i^1) = \overline{Y}_{ij}(t, \mathbf{x}_i^0) \). When \( \gamma < 0 \), note that \( \tilde{Y}_{ij}(t) \geq \tilde{Y}_{ij}(t) \) for all \( t \) if and only if \( \tilde{T}_{i1} \leq \tilde{T}_{i0} \). Suppose without loss of generality that in the coupled processes, subjects are relabeled in order of their infection in the constructed process, so the \( j \)th infection occurs in subject \( j, v_j = j \). Likewise the \( k \)th infection occurs in subject \( k, v_k = k \). Two cases are of interest. First, when \( j < k \) we have \( x_{il}^1 = x_{il}^0 \) for every \( l \leq j \), and so \( F_i(w) = G_i(w) \) for \( l \leq j < k \) and all \( w \). Therefore,

\[
\tilde{T}_{ij} = \sum_{l=1}^{j} \tilde{W}_{ij} = \sum_{l=1}^{j} F_i^{-1}(U_l) = \sum_{l=1}^{j} G_i^{-1}(U_l) = \sum_{l=1}^{j} \tilde{W}_{il} = \tilde{T}_{ij}. \quad (7)
\]
Second, when subject $k$ is infected first, or $k < j$, we have $F_l(w) = G_l(w)$ for $l < k$. However, for subjects $r$ infected after $k$ ($r > k$), we have

$$F_r(w) = 1 - \exp \left\{ -w \sum_{a \in S_r} e^{\eta_a} \left( \alpha + \sum_{b \in I_r} e^{\gamma x_{ib}^1 + \xi_{ib}} \right) \right\}$$

$$= 1 - \exp \left\{ -w \sum_{a \in S_r} e^{\eta_a} \left( \alpha + \sum_{b \in I_r, b \neq k} e^{\gamma x_{ib}^0 + \xi_{ib}} + e^{\gamma + \xi_{ik}} \right) \right\}$$

$$> 1 - \exp \left\{ -w \sum_{a \in S_r} e^{\eta_a} \left( \alpha + \sum_{b \in I_r, b \neq k} e^{\gamma x_{ib}^0 + \xi_{ib}} + e^{\gamma + \xi_{ik}} \right) \right\}$$

$$= 1 - \exp \left\{ -w \sum_{a \in S_r} e^{\eta_a} \left( \alpha + \sum_{b \in I_r} e^{\gamma x_{ib}^0 + \xi_{ib}} \right) \right\}$$

$$= G_r(w)$$

for all $w$. Therefore $F_{r}^{-1}(U_r) < G_{r}^{-1}(U_r)$ by monotonicity of $F_r(w)$ and $G_r(w)$, so the constructed infection times are

$$\tilde{T}_{ij}^1 = \sum_{l=1}^{j} \tilde{W}_{i}^1 = \sum_{l=1}^{j} F_{l}^{-1}(U_l) = \sum_{l=1}^{k} G_{l}^{-1}(U_l) + \sum_{r=k+1}^{j} F_{r}^{-1}(U_r)$$

$$< \sum_{l=1}^{k} G_{l}^{-1}(U_l) + \sum_{r=k+1}^{j} G_{r}^{-1}(U_r) = \sum_{l=1}^{j} \tilde{W}_{i}^0 = \tilde{T}_{ij}^0.$$
combinatorial identity

\[
\binom{n_i - 1}{m_i} = \frac{n_i - m_i}{m_i} \binom{n_i - 1}{m_i - 1},
\]

(10)

we can decompose a sum over allocations of \(m_i\) treatments to \(n_i - 1\) subjects into a sum over allocations of \(m_i - 1\) treatments to \(n_i - 1\) subjects, and an additional allocation of treatment to one more,

\[
\sum_{w \in \mathcal{P}_{m_i}^{-1}} \bar{Y}_{ij}(t, 0, w) = \frac{1}{m_i} \sum_{z \in \mathcal{P}_{m_i}^{-1}} \sum_{w \in \mathcal{P}_i(z)} \bar{Y}_{ij}(t, 0, w).
\]

(11)

The factor \(1/m_i\) appears in the right-hand side above because there are \(m_i\) allocations \(z\) for which a given \(w \in \mathcal{P}_i(z)\) is compatible; the double sum over-counts allocations by a factor of \(m_i\). Using this fact, we expand \(DE_{ij}(T)\) into a sum over allocations to subjects other than \(j\),

\[
DE_{ij}(T) = \binom{n_i - 1}{m_i - 1}^{-1} \sum_{z \in \mathcal{P}_{m_i}^{-1}} \bar{Y}_{ij}(T, 1, z) - \binom{n_i - 1}{m_i}^{-1} \sum_{w \in \mathcal{P}_{m_i}^{-1}} \bar{Y}_{ij}(T, 0, w)
\]

\[
= \binom{n_i - 1}{m_i - 1}^{-1} \left( \sum_{z \in \mathcal{P}_{m_i}^{-1}} \bar{Y}_{ij}(T, 1, z) - m_i \sum_{w \in \mathcal{P}_{m_i}^{-1}} \bar{Y}_{ij}(T, 0, w) \right)
\]

\[
= \binom{n_i - 1}{m_i - 1}^{-1} \frac{1}{n_i - m_i} \sum_{z \in \mathcal{P}_{m_i}^{-1}} (n_i - m_i) \bar{Y}_{ij}(T, 1, z) - \sum_{w \in \mathcal{P}_i(z)} \bar{Y}_{ij}(T, 0, w)
\]

\[
= \binom{n_i - 1}{m_i - 1}^{-1} \frac{1}{n_i - m_i} \sum_{z \in \mathcal{P}_{m_i}^{-1}} \sum_{w \in \mathcal{P}_i(z)} (\bar{Y}_{ij}(T, 1, z) - \bar{Y}_{ij}(T, 0, w))
\]

(12)

where the first equality follows from (2) under block randomization with \(m_i\) of \(n_i\) subjects treated, the second by (10), the third by (11), and the fourth because there are \(n_i - m_i\) terms in the sum over \(w \in \mathcal{P}_i(z)\).

Therefore, \(DE_{ij}(T)\) can be expressed as a sum of contrasts between the average outcome of \(j\) under joint treatments \((1, z)\) and \((0, w)\) where \(w\) is the same as \(z\), but with one additional treated subject. Each contrast in the last line of (12) has sign as given above, and the result follows.

The proof of Proposition 3 is very similar and is presented in the Supplement. Three final results generalize the results for the individual average direct effect \(DE_{ij}(T)\) to the cluster and population average direct effect estimands. The proofs, which rely only on Propositions 1, 2, and 3 and the definitions of \(DE_i(T)\) and \(DE(T)\), are omitted.
Corollary 1 (Cluster and population average DE under Bernoulli randomization) Suppose $\beta = 0$ and treatment assignment is Bernoulli randomized. Then $DE_i(T) = DE(T) = 0$.

Corollary 2 (Cluster and population average DE under block randomization) Suppose $\beta = 0$, treatment assignment is block randomized. If $\gamma < 0$ then $DE_i(T) > 0$ and $DE(T) > 0$; if $\gamma = 0$ then $DE_i(T) = DE(T) = 0$; and if $\gamma > 0$ then $DE_i(T) < 0$ and $DE(T) < 0$.

Corollary 3 (Cluster and population average DE under cluster randomization) Suppose $\beta = 0$ and treatment assignment is cluster randomized. If $\gamma < 0$ then $DE_i(T) < 0$ and $DE(T) < 0$; if $\gamma = 0$ then $DE_i(T) = DE(T) = 0$; and if $\gamma > 0$ then $DE_i(T) > 0$ and $DE(T) > 0$.

4.2 Simulation study

We investigate the properties of the population average direct effect as the true infectiousness effect $\gamma$ changes. The hazard of infection takes the form of (3) where the null hypothesis is $\beta = 0$ and we investigate $DE(T)$ as a function of $\gamma \in [-2, 2]$. The exogenous force of infection is $\alpha = 0.01$, the individual susceptibility coefficients $\eta_{ij}$ are independent Normal($\mu_\eta, \sigma^2_\eta$) and infectiousness coefficients $\xi_{ij}$ are independent Normal($\mu_\xi, \sigma^2_\xi$). Unless otherwise noted, the cluster size $n_i$ is $2 + \text{Poisson}(2)$, the observation time is $T = 10$, and all subjects were uninfected at baseline, $Y_{ij}(0) = 0$. The Supplement provides additional details about the simulation setting.

Figure 2 shows simulation results validating the analytic derivations above. Under Bernoulli randomization $DE(T)$ is zero for any $\gamma$; under block randomization it has the opposite sign as $\gamma$; and under cluster randomization it has the same sign as $\gamma$. Figure 3 shows properties of $DE(T)$ as a function of $\gamma$ under various epidemiologic and study design parameters, when $\beta = 0$. The top row shows results under block randomization, and the bottom row shows results under cluster randomization. The left column shows $DE(T)$ for increasing values of $\sigma^2$, the variability of individual-level susceptibility and infectiousness. The middle column shows how $DE(T)$ changes with $\mu_\xi$, the average value of the individual-level infectiousness coefficient. When these values are large and negative, few infections are transmitted by infected individuals, so the value of $\gamma$ has little effect on $DE(T)$, which stays near zero. When $\mu_\xi$ is large and positive, something similar happens: infected individuals are highly infectious even when $\gamma < 0$, so $DE(T)$ is near zero for a wide range of values of $\gamma$. When $\mu_\xi$ is near zero, the value of $\gamma$ fully determines the infectiousness of treated individuals, and $DE(T)$ exhibits the largest difference from zero. In the right column, we examine the effect of changes and heterogeneity in the follow-up time $T$, allowing the observation time $T_i$ to vary between clusters. In all cases, the magnitude of the direct effect increases with the absolute value of $\gamma$. While Propositions 1–3 give the sign of $DE(T)$ for any combination of parameter values, simulation results show that the magnitude of $DE(T)$ changes substantially depending on the specific study design and epidemiologic characteristics. In the Supplement we present a simulation study exploring the properties of $DE(T)$ when $\beta \neq 0$. “Springer
Fig. 2 Simulation results for $DE(T)$ under the null hypothesis of no susceptibility effect $\beta = 0$, as a function of the infectiousness effect $\gamma$, for different randomization designs and cluster size distributions. Bernoulli randomization (black) recovers $DE(T) = 0$ for any $\gamma$. Block randomization (red) shows $DE(T)$ has the opposite sign as $\gamma$, and cluster randomization (blue) shows $DE(T)$ has the same sign as $\gamma$ (color figure online)

5 Discussion

Greenwood and Yule (1915) proposed three conditions for making valid inferences about the effect of a vaccine: (1) “The persons must be, in all material respects, alike”; (2) “The effective exposure to the disease must be identical in the case of inoculated and uninoculated persons”; and (3) “The criteria of the fact of inoculation and of the fact of the disease having occurred must be independent”. Randomization ensures that conditions 1 and 3 are satisfied on average (Rothman et al. 2008; Greenland and Robins 1986). The direct effect defined by Hudgens and Halloran (2008) compares individual infection outcomes in a way that ensures condition 2 does not hold: treated and untreated subjects experience differential exposure to infectiousness, and $DE_{ij}(t)$ is subject to confounding.
The direct effect is a well-defined and natural statistical estimand that is identified under randomization with mild assumptions. But under some randomization designs, it may not provide empirical researchers with the individual causal effect they seek, because it does not hold all other things equal. A heuristic explanation provides useful intuition.

1. Under Bernoulli randomization, treated and untreated subjects are exposed to the same number of treated individuals on average.
2. Under block randomization, treated subjects are exposed to fewer treated individuals \((m_i - 1)\) than untreated subjects \((m_i)\).
3. Under cluster randomization, treated subjects are exposed to more treated individuals \((n_i - 1)\) than untreated subjects \((0)\).

These differences in joint treatment distribution are natural consequences of the randomization designs; Propositions 1–3 establish the connection to differential exposure to infectiousness, and to the direct effect estimand under the structural transmission model (3). When the null hypothesis of \(\beta = 0\) is true and an infectiousness effect exists \((\gamma \neq 0)\), treated and untreated subjects under block and cluster randomization experience differential exposure to infectiousness that depends on the sign of the infectiousness coefficient \(\gamma\). These results apply to individuals within clusters, and hold for any number of clusters. Similarly, odds and risk ratios computed by contrasting
average individual outcomes under treatment versus no treatment may be subject to
the same biases (e.g. Morozova et al. 2018). Notably, Karwa and Airoldi (2018) find
similar results, characterizing the bias of the direct effect with a linear outcome model
and interference under block randomization and Bernoulli designs.

Our main results investigate the direct effect under the null hypothesis \( \beta = 0 \)
because this case is analytically tractable, and because preservation of the null is
a desirable property of any effect measure or test statistic. Some real-world interven-
tions may have this feature; for example, transmission-blocking vaccines (Kaslow
2002) have negligible susceptibility effect, but may be effective in reducing infectious-
ness of infected individuals. Isolation policies may also confer minimal susceptibility
benefit to individuals assigned to “quarantine upon infection”, and a strong beneficial
infectiousness effect on their contacts (Aiello et al. 2016). For untested interventions
like new vaccines, investigators may not know whether the susceptibility effect is
beneficial, harmful, or null. The results outlined here may apply in cases where the
true susceptibility effect \( \beta \) is nonzero: when the average infection outcome \( \bar{Y}_{ij}(t, x_i) \)
is a continuous function of \( \beta \), there may exist an interval around \( \beta \neq 0 \) in which
the direct effect is biased across the null hypothesis of no susceptibility effect under
some designs. Indeed, prior work demonstrated analytically that in a simple 2-person
cluster case under block randomization the direct effect can have a sign opposite that
of a non-null susceptibility effect (Morozova et al. 2018; Cai et al. 2021).

Therefore estimating \( DE(t) = 0 \) under block randomization need not imply that
the susceptibility effect is null, nor does estimating \( DE(t) \neq 0 \) imply that the sus-
ceptibility effect is not null. In particular, simulation results show that under block
randomization, a vaccine that both helps prevent infection in each person who receives
it \( (\beta < 0) \) and helps prevent transmission upon infection \( (\gamma < 0) \) can nevertheless
exhibit \( DE(t) > 0 \). When \( DE(t) \) is interpreted as a causal parameter, investigators
may conclude that an effective intervention is harmful to the individuals who receive it
because its “direct effect” is positive. Simulation results in the Supplement explore the
conditions leading to sign mismatch between \( DE(t) \) and \( \beta \neq 0 \). These results are con-
sistent with analytic results for direction bias in the simple 2-person case (Morozova
et al. 2018).

In this paper, we employ a relatively simple structural transmission model (3)
because it is widely used and well understood by infectious disease epidemiologists,
the hazard of infection has a simple functional form, and its parameters \( \beta \) and \( \gamma \)
correspond naturally to the susceptibility and infectiousness effects defined by Hal-
loran et al. (1997). Mathematical transmission models are always a simplification
of reality (Britton and Lindenstrand 2009), but we believe the one employed here
captures many of the essential features of transmission that have implications for
the design of randomized vaccine trials. However, this transmission model does not
incorporate additional realistic features of infectious disease transmission, such as
latency, removal/recovery, changing infectiousness over time, or treatment following
infection. Similarly, we have not modeled heterogeneous contact patterns within clus-
ters, nor violated stratified interference by permitting transmission between clusters.
We conjecture that more sophisticated structural models of infectious disease trans-
mision would not differ in their qualitative implications: dependent randomization
designs induce differential exposure to infectiousness whenever the treatment affects
infectiousness, resulting in confounding of the direct effect as a measure of the susceptibility effect. If the direct effect under dependent randomization designs does not provide a meaningful approximation to the susceptibility effect of interest under a simplistic transmission model such as (3), we do not expect it to do so under a richer class of more complex structural transmission models.

Researchers who wish to avoid the pathologies of the direct effect in a randomized trial have three basic options. First, Proposition 1 shows that changing the randomization design to Bernoulli allocation within clusters breaks the dependence between \( x_{ij} \) and \( x_i^{(j)} \) (Sävje et al. 2021). Then the conditional probability \( \Pr(X_i^{(j)} = x_i^{(j)} | x_{ij} = x) \) in (2) becomes the marginal probability \( \Pr(X_i^{(j)} = x_i^{(j)}) \), and the direct effect becomes a simple average of individual effects. Second, researchers may target a marginal estimand that does not condition on the assigned treatment. This approach would permit use of a dependent randomization design by changing the conditional marginalizing distribution in (2) to the unconditional distribution of the treatment to other units, \( \Pr(X_i^{(j)} = x_i^{(j)}) \), provided this probability is positive under the design. Specifically, this expected average treatment effect (EATE) is a generalization of the direct effect proposed by Sävje et al. (2021) which marginalizes over all possible treatment assignments to other units under the randomization design. Construction of the EATE involves changing the estimand so that the marginalizing distribution (the randomization distribution of other units’ treatments) is invariant to the treatment assigned to individual \( j \), and is therefore the same on both sides of the causal contrast. Third, when structural assumptions are warranted and enough data are available, researchers may choose to fit a structural model similar to (3) to estimate parameters (e.g. \( \beta \)) corresponding to the causal effects of interest (Auranen et al. 2000; Cauchemez et al. 2006; Kenah 2014).

Finally, we have focused here on three idealized randomization designs that are employed in real-world intervention trials. Non-randomized (i.e. pragmatic, or observational) studies of interventions or risk factors for infection in clusters occupy an uncertain middle ground. Even when the intervention or covariate of interest is unrelated to other baseline confounders and independent of the potential infection outcomes, it may be unreasonable to assume that it is distributed independently at random within clusters, as it would be under Bernoulli randomization. Likewise, strict negative or positive correlation in covariate values, of the kind induced by block and cluster randomization respectively, seems implausible. When any dependence exists in the distribution of treatment in an observational study, regression adjustment or stratification on baseline covariates may not be sufficient to ensure exchangeability of subjects with respect to infection exposure during the study. Depending on the distribution of treatment, the relationship between the direction or sign of marginal contrasts and the true susceptibility effect may be difficult to predict.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00285-022-01801-8.

Acknowledgements We are especially grateful to Michael Hudgens and M. Elizabeth Halloran for helpful comments. In addition, we thank P. M. Aronow, Xiaoxuan Cai, Ted Cohen, Soheil Eshghi, Gregg S. Gonsalves, Eben Kenah, Zehang Li, Wen Wei Loh, Sida Peng, Fredrik Sävje, Yushuf Sharker, and Daniel...
Weinberger for helpful comments. This work was supported by NIH grants NICHD DP2 HD091799-01 and NIDA R36 DA042643.

**Funding** This work was supported by NIH grants NICHD DP2 HD091799-0 and NIDA R36 DA042643.

**Data availability** All data and code are available upon request.

**Declarations**

**Conflict of interest** The authors claim no conflicts of interest.

**References**

Acosta CJ, Galindo CM, Mohammad A, Abu ER, Leon OR, Carolina D-HM, Anne-Laure P, Dinh TV, Yang J, Kyung PJ et al (2005) A multi-country cluster randomized controlled effectiveness evaluation to accelerate the introduction of Vi polysaccharide typhoid vaccine in developing countries in Asia: rationale and design. Trop Med Int Health 10(12):1219–1228

Aiello AE, Simanek AM, Eisenberg MC, Walsh AR, Davis B, Volz E, Cheng C, Rainey JJ, Uzicanin A, Gao H et al (2016) Design and methods of a social network isolation study for reducing respiratory infection transmission: the e-X-FLU cluster randomized trial. Epidemics 15:38–55

Anderson RM, May RM (1992) Infectious disease of humans, dynamics and control. Oxford University Press, Oxford

Andersson H, Britton T (2000) Stochastic epidemic models and their statistical analysis. Springer, New York

Auranen K, Arjas E, Leino T, Takala AK (2000) Transmission of pneumococcal carriage in families: a latent Markov process model for binary longitudinal data. J Am Stat Assoc 95:1044–1053

Becker Niels G (1989) Analysis of infectious disease data. Chapman and Hall London, London

Becker NG, Britton T (2004) Estimating vaccine efficacy from small outbreaks. Biometrika 91(2):363–382

Becker NG, Britton T, O’Neill PD (2003) Estimating vaccine effects on transmission of infection from household outbreak data. Biometrics 59(3):467–475

Becker NG, Britton T, O’Neill PD (2006) Estimating vaccine effects from studies of outbreaks in household pairs. Stat Med 25(6):1079–1093

Belshe RB, Mendelman PM, Trenar J, King J, Gruber WC, Piedra P, Bernstein DI, Hayden FG, Kotloff K, Zangwill K et al (1998) The efficacy of live attenuated, cold-adapted, trivalent, intranasal influenza virus vaccine in children. New England J Med 338(20):1405–1412

Britton T, Lindenstrand D (2009) Epidemic modelling: aspects where stochasticity matters. Math Biosci 222(2):109–116

Buchanan AL, Vermund SH, Friedman SR, Spiegelman D (2018) Assessing individual and disseminated effects in network-randomized studies. Am J Epidemiol 187(11):2449–2459

Cai X, Loh WW, Crawford FW (2021) Identification of causal intervention effects under contagion. J Causal Inference 9(1):9–38

Cauchemez S, Carrat F, Viboud C, Valleron AJ, Boëlle PY (2004) A Bayesian MCMC approach to study transmission of influenza: application to household longitudinal data. Stat Med 23:3469–87

Cauchemez S, Temime L, Guillemot D, Varon E, Valleron A-J, Thomas G, Boëlle P-Y (2006) Investigating heterogeneity in pneumococcal transmission: a Bayesian MCMC approach applied to a follow-up of schools. J Am Stat Assoc 101(475):946–958

Cauchemez S, Donnelly CA, Reed C, Ghani AC, Fraser C, Kent CK, Finelli L, Ferguson NM (2009) Household transmission of 2009 pandemic influenza A (H1N1) virus in the United States. N Engl J Med 361(27):2619–2627

Cole SR, Lau B, Eron JJ, Brookhart MA, Kitahata MM, Martin JN, Mathews WC, Mugavero MJ (2015) Estimation of the standardized risk difference and ratio in a competing risks framework: application to injection drug use and progression to AIDS after initiation of antiretroviral therapy. Am J Epidemiol 181(4):238–245

© Springer
den Hollander F (2012) Probability theory: the coupling method. Mathematical Institute, Leiden University. URL https://urldefense.com/v3/http://websites.math.leidenuniv.nl/probability/lecturenotes/CouplingLectures.pdf

Devroye L (1986) Non-uniform random variate generation. Springer, Cham

Diekmann O, Heesterbeek H, Britton T (2012) Mathematical tools for understanding infectious disease dynamics. Princeton University Press, Princeton

Greenland S, Robins JM (1986) Identifiability, exchangeability, and epidemiological confounding. Int J Epidemiol 15(3):413–419

Greenwood M, Yule GU (1915) The statistics of anti-typhoid and anti-cholera inoculations, and the interpretation of such statistics in general

Elizabeth Halloran M, Struchiner CJ (1995) Causal inference in infectious diseases. Epidemiology 6:142–151

Halloran ME, Haber M, Longini IM, Struchiner CJ (1991) Direct and indirect effects in vaccine efficacy and effectiveness. Am J Epidemiol 133(4):323–331

Elizabeth Halloran M, Struchiner Claudio J, Longini IM Jr (1997) Study designs for evaluating different efficacy and effectiveness aspects of vaccines. Am J Epidemiol 146(10):789–803

Elizabeth Halloran M, Longini Ira M, Struchiner Claudio J (2010) Design and analysis of vaccine studies. Springer, Cham

Hayden FG, Gubareva LV, Monto AS, Klein TC, Elliott MJ, Hammond JM, Sharp SJ, Ossi MJ (2000) Inhaled zanamivir for the prevention of influenza in families. New England J Med 343(18):1282–1289

Hudgens MG, Halloran ME (2008) Toward causal inference with interference. J Am Stat Assoc 103(482):832–842

Karwa V, Airoldi EM (2018) A systematic investigation of classical causal inference strategies under misspecification due to network interference. arXiv preprint arXiv:1810.08259

Kaslow DC (2002) Transmission-blocking vaccines. Malaria Immunol 80:287–307

Kenah E (2013) Non-parametric survival analysis of infectious disease data. J R Stat Soc B 75(2):277–303

Kenah E (2014) Semiparametric relative-risk regression for infectious disease transmission data. J Am Stat Assoc 110:313–325

Longini IM Jr, Hudgens MG, Halloran ME, Sagatelian K (1999) A Markov model for measuring vaccine efficacy for both susceptibility to infection and reduction in infectiousness for prophylactic HIV vaccines. Stat Med 18(1):53–68

Monto AS, Pichichero ME, Blanckenberg SJ, Ruuskanen O, Cooper C, Fleming DM, Kerr C (2002) Zanamivir prophylaxis: an effective strategy for the prevention of influenza types A and B within households. J Infect Dis 186(11):1582–1588

Morozova O, Cohen T, Crawford FW (2018) Risk ratios for contagious outcomes. J R Soc Interface 15(138):20170696

O’Hagan JJ, Lipsitch M, Hernán MA (2014) Estimating the per-exposure effect of infectious disease interventions. Epidemiology 25(1):134–138

O’Neill PD, Balding DJ, Becker NG, Eerola M, Mollison D (2000) Analyses of infectious disease data from household outbreaks by Markov chain Monte Carlo methods. J R Stat Soc Ser C 49(4):517–542

Rhodes PH, Halloran ME, Longini IM Jr (1996) Counting process models for infectious disease data: distinguishing exposure to infection from susceptibility. J R Stat Soc B 58:751–762

Ross SM (1996) Stochastic processes. Wiley, New York

Rothman KJ, Greenland S, Lash TL (2008) Modern epidemiology. Lippincott Williams & Wilkins, Philadelphia

Rubin DB (2005) Causal inference using potential outcomes: design, modeling, decisions. J Am Stat Assoc 100(469):322–331

Sävje F, Aronow PM, Hudgens MG (2021) Average treatment effects in the presence of unknown interference. Ann Stat 49(2):673–701

Simondon F, Preziosi M-P, Yam A, Kane CT, Chabirand L, Iteman I, Sanden G, Mboup S, Hoffenbach A, Knudsen K et al (1997) A randomized double-blind trial comparing a two-component acellular to a whole-cell pertussis vaccine in Senegal. Vaccine 15(15):1606–1612

Sobel ME (2006) What do randomized studies of housing mobility demonstrate? Causal inference in the face of interference. J Am Stat Assoc 101(476):1398–1407

Struchiner CJ, Halloran ME (2007) Randomization and baseline transmission in vaccine field trials. Epidemiol Infect 135(2):181–194
Tsang TK, Cowling BJ, Fang VJ, Chan K-H, Dennis KM, Leung GM, Peiris JSM, Cauchemez S (2015) Influenza A virus shedding and infectivity in households. J Infect Dis 212(9):1420–1428
Tsang TK, Fang VJ, Chan K-H, Dennis KM, Leung GM, Peiris JSM, Cowling BJ, Cauchemez S (2016) Individual correlates of infectivity of influenza A virus infections in households. PloS One 11(5):e0154418
van Boven M, Ruijs WLM, Wallinga J, O’Neill PD, Hahne S (2013) Estimation of vaccine efficacy and critical vaccination coverage in partially observed outbreaks. PLoS Comput Biol 9(5):e1003061
Vander Weele TJ, Robins JM (2012) Stochastic counterfactuals and stochastic sufficient causes. Stat Sin 22(1):379
Vander Weele TJ, Tchetgen EJT (2011) Effect partitioning under interference in two-stage randomized vaccine trials. Stat Prob Lett 81(7):861–869
Welliver R, Monto AS, Carewicz O, Schatteman E, Hassman M, Hedrick J, Jackson HC, Huson L, Ward P, Oxford JS et al (2001) Effectiveness of oseltamivir in preventing influenza in household contacts: a randomized controlled trial. J Am Med Assoc 285(6):748–754

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.