HHS Public Access

Author manuscript

J Causal Inference. Author manuscript; available in PMC 2021 October 20.

Published in final edited form as:

J Causal Inference. 2021 January ; 9(1): 9–38. doi:10.1515/jci-2019-0033.

Identification of causal intervention effects under contagion

Xiaoxuan Cai[1], **Wen Wei Loh**[2], **Forrest W. Crawford**[1,3,4,5]

[1]Department of Biostatistics, Yale School of Public Health

[2]Department of Data Analysis, University of Ghent

[3]Department of Statistics & Data Science, Yale University

[4]Department of Ecology and Evolutionary Biology, Yale University

[5]Yale School of Management

Abstract

Defining and identifying causal intervention effects for transmissible infectious disease outcomes is challenging because a treatment – such as a vaccine – given to one individual may affect the infection outcomes of others. Epidemiologists have proposed causal estimands to quantify effects of interventions under contagion using a two-person partnership model. These simple conceptual models have helped researchers develop causal estimands relevant to clinical evaluation of vaccine effects. However, many of these partnership models are formulated under structural assumptions that preclude realistic infectious disease transmission dynamics, limiting their conceptual usefulness in defining and identifying causal treatment effects in empirical intervention trials. In this paper, we propose causal intervention effects in two-person partnerships under arbitrary infectious disease transmission dynamics, and give nonparametric identification results showing how effects can be estimated in empirical trials using time-to-infection or binary outcome data. The key insight is that contagion is a causal phenomenon that induces conditional independencies on infection outcomes that can be exploited for the identification of clinically meaningful causal estimands. These new estimands are compared to existing quantities, and results are illustrated using a realistic simulation of an HIV vaccine trial.

Keywords

infectiousness; interference; mediation; susceptibility; transmission; vaccine

1 Introduction

Estimating the causal effect of an intervention can be challenging when the outcome of interest is contagious [41]. For example, a vaccine intended to prevent infection by a transmissible disease may reduce the risk of infection in individuals who receive it, and may reduce transmissibility if a vaccinated individual becomes infected. When study subjects are embedded in interacting groups among whom the disease may be transmitted, it can be difficult to separate the effect of one subject's vaccination on themselves from its effect on other individuals and the group as a whole. Usually, the estimand of greatest clinical interest is the effect of an intervention on individual risks of infection, holding all else constant.

The pursuit of empirically meaningful definitions of population-level causal vaccine effects has a long history. Greenwood and Yule [19] first described informally the conditions under which vaccine effects can be estimated. Halloran et al. [24] established some of the first theory and definitions for clinically meaningful vaccine effects, and subsequent work by Halloran and colleagues [22, 26, 27] described epidemiological study designs for identifying these quantities. Halloran and Struchiner [23] gave the first formal definitions of causal vaccine estimands using notation and assumptions of a modern counterfactual-based causal inference framework [54]. Hudgens and Halloran [31] and Tchetgen Tchetgen and VanderWeele [60] showed how this formalism could be applied in empirical randomized trials of clustered individuals [21, 29]. More recently, researchers have shown that randomized trials may not measure clinically meaningful intervention effects when infection can be transmitted within groups [15, 40, 59].

Researchers have described two-person partnership models of infectious disease transmission for defining more granular, or individual, causal intervention effects. VanderWeele and Tchetgen Tchetgen [64] introduced a partnership model consisting of two interacting individuals who may be vaccinated and can transmit the infection to each other. By limiting the extent of potential disease transmission to two individuals, effects can be more easily defined in terms of potential outcomes indexed by treatments of both individuals and the outcome of their partner. The partnership model can accommodate many types of epidemiological relationship where infectious disease transmission may occur between indidivudals. The partnership model can accommodate, for example, parent-child relationships, sibling relationships, needle-sharing partnerships among injection drug users, or sexual partnerships. While nearly all real-world partnerships occur in the context of a broader network of epidemiological relationships with others, partnership models may be useful when pairs are drawn nearly independently from disparate networks, so that pairs experience independent exposure to infection from outside the partnership. For example, a study of disease transmission among cohabitating couples chosen from different cities could plausbly claim that the pairs experienced independent exposure to infection from outside the relationship.

Using a principal stratification approach, the partnership model permits computations of bounds for the infectiousness effect [10, 20, 64]. VanderWeele et al. [66] presented a special case of the partnership model in which one individual is home-bound, and can only be infected via transmission from the other. The assumed asymmetry in the disease transmission structure – the home-bound partner cannot be infected from a source external to the partnership and cannot infect the other partner – makes this model tractable for point identification of contagion and infectiousness effects by ensuring that *interference* only happens in one direction. Interference arises when an individual's potential outcomes depend on the treatment status of others [13]. To allow for mutual dependence of individuals' potential outcomes on others' treatments, Ogburn and VanderWeele [43] extend this approach to allow both individuals to be treated, with transmission occurring only from one specified individual to the other. However, Ogburn and VanderWeele [42] show using causal diagrams that transmission complicates application of existing mediation techniques, requiring additional structural assumptions about the nature of dependence among outcomes under different forms of interference [6, 44, 55, 57]. Shpitser et al. [57] proposed extensions

of mediation analysis to symmetric mediation settings, using statistical chain graph models that do not require a priori fixing the individual whose outcome plays the role of the mediator within the partnership.

When the outcomes are time-dependent processes – as is often in infectious disease transmission dynamics – binary outcome indicators and specified time windows may be used to define outcomes so that the mediation-based approaches may be applied. But these definitions can complicate identification of causal effects because (i) a repeatedly measured outcome over time may introduce multiple mediators, and (ii) absence of the outcome at prior time points as a prerequisite for later measurements induces time-varying confounding. Existing methods for longitudinal mediation analysis have therefore either focused on defining "interventional" indirect effects in terms of combined path-specific effects that can be non-parametrically identified [36, 56, 65, 67, 69], or adopted approaches that avoid defining nested counterfactuals for time-to-event outcomes [1, 14]. These approaches to longitudinal mediation share the common prerequisite that the roles of the outcomes within each partnership are asymmetric.

Statisticians and epidemiologists have developed parallel literature devoted to mathematical modeling of infectious disease transmission dynamics. This work treats infectious disease transmission as a dynamic temporal phenomenon: the risk of infection in a given subject may change over time, as a function of the infection status of their contacts, and covariates. For example, Rhodes et al. [51] present hazard models of infectious disease transmission in groups that accommodate individual-level (e.g. treatment) variables with possibly different effects on susceptibility and infectiousness. Kenah [33, 34] extends these ideas to develop nonparametric and semi-parametric statistical models for estimating covariate effects under contagion. Structural transmission modeling has gained wide use in clinical studies of infectious disease dynamics because it combines mechanistic assumptions about infectious disease transmission with regression-style covariate adjustment [3–5, 7–9, 46, 61, 62, 68].

In this paper, we take a different approach to define and identify intervention effects in symmetric two-person partnerships under contagion. We seek to combine approaches from causal mediation analysis and mathematical modeling of transmission to develop a nonparametric framework that formalizes the role of time in infectious disease transmission from a causal perspective. In our construction, either individual can be vaccinated, can be infected from outside, and can infect the other if infected themselves. An individual's treatment (or vaccine status) and covariates may affect both susceptibility to, and infectiousness of, their infection outcome. We first introduce a generic causal model and straightforward assumptions that permit non-parametric identification of "exposure-controlled" and natural "exposure-marginalized" contagion, susceptibility, and infectiousness effects. Briefly, the contagion effect captures how transmissible the infection is from an infected individual to an uninfected individual. The susceptibility effect summarizes the effect of treatment on the infection outcome of individual who receives it. The infectiousness effect indicates the effect of an individual's treatment on others' outcomes, when that individual is infected. We propose a framework that is non-parametric and imposes no restrictions on the joint distribution of infection times in a partnership. Before any infections have occurred in a partnership, the potential first infection times are

conditionally independent, because neither partner can yet transmit the infection to the other. After the first infection, the time to infection of the remaining susceptible partner is now a function their partner's, as well as their own, treatment and covariates. Because the resulting causal model incorporates this temporally changing structure, it is more complex than settings considered in other proposals. In particular, the causal effects defined in this paper differ from the "direct" and "indirect" effects defined using the interference framework developed by [31] in ways we describe formally. On the other hand, this added complexity yields straightforward point identification results that cannot be obtained by treating the infection outcomes of both individuals as simultaneous mediating variables [10, 20, 64]. Lastly, we discuss nonparametric identification under randomization and in observational settings, compare these estimands to existing quantities proposed by other authors, and conduct a simulation analysis of a hypothetical HIV vaccine trial to illustrate the estimands.

2 Setting

Consider a population consisting of pairs of individuals, henceforth referred to as partnerships. Within a partnership, either individual can be infected from an external source (exogenous to the partnership), and once infected, an individual may internally (endogenous to the partnership) transmit the infection to their uninfected partner. Label the individuals in the partnership 1 and 2. In a given partnership, let T_i be the infection time of person i, and let $Y_i(t) = \mathbf{1}\{T_i < t\}$ be the indicator of prior infection at time t. Let X_i be the binary treatment status of i, and let $\mathbf{X} = (X_1, X_2)$ be the joint binary treatment vector for the partnership. Let $\mathbf{L} = (\mathbf{L}_1, \mathbf{L}_2)$ be measured baseline covariates for the two individuals, including shared covariates for the partnership as a whole. In each partnership, we observe $(T_1, T_2, X_1, X_2, \mathbf{L}_1, \mathbf{L}_2)$. In a symmetric partnership, the labels for individuals 1 and 2 may carry meaning (e.g. in mother-child pairs), or may be arbitrary and interchangeable. We will use the index i to refer generically to one individual, either 1 or 2, and j to refer to the partner of i.

To describe the causal structure of infectious disease transmission within a partnership, we consider a decomposition of the infection time T_i that will help us define counterfactual infection times under different circumstances. Recall that both individuals are uninfected at baseline, and let W_i be the time to initial infection of i from a source of infection external to the partnership. If i is the first in their partnership to become infected, then we observe W_i. If their partner j is infected first, we observe $W_j = w_j$ and W_i is *censored* at time w_j. When W_i is censored by earlier infection of j, let Z_i be the additional time to infection of i, beyond the infection time w_j of their partner. Formally, we decompose T_i as follows.

$$T_i = \begin{cases} W_i & \text{if } W_i < W_j \\ W_j + Z_i & \text{otherwise.} \end{cases} \tag{1}$$

We emphasize that the decomposition (1) is purely notational, and places no *a priori* restrictions on the joint distribution of infection times (T_i, T_j). Instead, (1) shows how observation of (T_i, T_j) reveals information about these infection waiting times: if $T_i < T_j$, then we can determine $W_i = T_i$, $W_j > T_i$, and $Z_j = T_j - T_i$. Figure 1 illustrates this decomposition and motivates the contagion effect presented in Definition 1 below: the

disease is said to be "contagious" if the distribution of T_i is different from that of W_i, or equivalently, if prior infection of j ($W_j < W_i$) changes the conditional distribution of the remaining time to infection of i (Z_i). The definition (1) permits specification of causal assumptions, outlined below, to capture the way treatments to both i and j may affect different parts of the waiting times to infection.

In line with existing partnership models, it is assumed throughout that the partnerships are independent, thereby ruling out transmission between partnerships [31, 53]. Though partnerships are assumed to be independent, the waiting times W_i and W_j, or Z_i and Z_j, need not be identically distributed. The potential for transmission between partners is assumed to be symmetric – that is, either can infect the other – but the framework accommodates asymmetries in transmission if the distributions of W_i and W_j, or Z_i and Z_j, differ.

2.1 Assumptions

In this section, we describe assumptions that are sufficient to identify the causal effects defined in Section 3 from observable infection time data for each partnership. We state assumptions for a generic individual i and their partner j. To define potential, or counterfactual, infection times for individual i, let $W_i(\mathbf{x})$ be the potential value of W_i under the joint treatment allocation $\mathbf{x} = (x_1, x_2)$. Let $Z_i(w_j, \mathbf{x})$ be the additional potential time to infection of i, following the infection of j at time $W_j(\mathbf{x}) = w_j$, under joint treatment allocation $\mathbf{x}$.

Assumption 1 (Exclusion restriction and independence of the initial infection).— $W_i(\mathbf{x}) = W_i(x_i), W_i(x_i) \perp\!\!\!\perp W_j(x_j) | \mathbf{L}$, *and* $W_i(x_i) \perp\!\!\!\perp \mathbf{L}_j | \mathbf{L}_i$, *for all* $\mathbf{x}$.

Assumption 1 states that individual i's initial infection time $W_i(\mathbf{x})$ is invariant to the partner's treatment status x_j. Hence it may be viewed as a "no-interference" assumption on W_i, because W_i is the initial infection time from an external source, which can only be realized when W_i precedes W_j. Further, $W_i(x_i)$ is independent of $W_j(x_j)$ given (observed) covariates $\mathbf{L}$. Assumption 1 respects a unique property of infectious disease: neither transmission nor treatment interference can occur without prior infection.

Assumption 2 (Initial infection exchangeability).—
$Z_i(w_j, \mathbf{x}) \perp\!\!\!\perp W_j(x_j) | \mathbf{L}$, *for all* $\mathbf{x}$, $w_j > 0$.

Assumption 2 states that there is sufficient covariate information in $\mathbf{L}$ so that the potential further time to infection $Z_i(w_j, \mathbf{x})$ when j is infected at w_j is conditionally independent of the potential initial infection time $W_j(x_j)$ of j. While this assumption bears similarity to the assumption of no unobserved confounding between the counterfactual mediator and nested potential outcome (for the same individual) under the single mediator setting [49, 52], we note that this assumption relates to counterfactual outcomes for different individuals.

Assumption 3 (Treatment exchangeability).—
$W_i(x_i) \perp\!\!\!\perp \mathbf{X} | \mathbf{L}$, *and* $Z_i(w_j, \mathbf{x}) \perp\!\!\!\perp \mathbf{X} | \mathbf{L}$, *for all* $\mathbf{x}$, $w_j > 0$.

Assumption 3 means that the potential waiting times $W_i(x_i)$ and $Z_i(w_j, \mathbf{x})$ are independent of the assigned treatment $\mathbf{X}$ within levels of the (observed) covariates $\mathbf{L}$. This assumption *prima facie* resembles the conventional unconfoundedness assumptions for the (individual) exposure-mediator and exposure-outcome relations in mediation analysis. But in this context, Assumption 3 states that there is no unmeasured confounding between an individual's infection times and the joint treatments for both individuals in the same partnership.

Two additional assumptions, commonly made in the literature when identifying causal estimands, ensure identifiability of potential infection outcomes from observational data.

Assumption 4 (Consistency).— $W_i = W_i(x_i)$, *and* $Z_i = Z_i(w_j, \mathbf{x})$ *under the observed treatment* $\mathbf{X} = \mathbf{x}$ *and infection time* $W_j = w_j$, *for all* $\mathbf{x}$, $w_j > 0$.

Assumption 5 (Positivity).— $0 < \Pr(W_j < w | X_i = x_i, \mathbf{L}_i = \mathbf{l}_i) < 1$ *for all* $w > 0$, x_i, *and* $\mathbf{l}_i$; $0 < \Pr(Z_j < z | \mathbf{X} = \mathbf{x}, \mathbf{L} = \mathbf{l}) < 1$ *for all* $z > 0$, $\mathbf{x}$ *and* $\mathbf{l}$; *and* $0 < \Pr(\mathbf{X} = \mathbf{x} | \mathbf{L} = \mathbf{l}) < 1$ *for all* $\mathbf{l}$.

A final assumption permits identification of certain "cross-world" potential infection outcomes.

Assumption 6 (Cross-world initial infection exchangeability).—
$Z_i(w_j, \mathbf{x}) \perp\!\!\!\perp W_j(x'_j) \mid \mathbf{L}$ *when* $\mathbf{x} = (x_i, x_j)$ *and* $x'_j \neq x_j$, *for all* w_j, $\mathbf{x}$, *and* x'_j.

Assumption 6 states that within levels of the observed covariates $\mathbf{L}$, the potential waiting time of i to infection, after j is infected at w_j under treatment x_j, is independent of the potential infection time W_j under a different treatment x'_j. Informally, when Assumption 6 holds, after j becomes infected at (some fixed time) w_j, the waiting time until i becomes infected under treatments $\mathbf{x} = (x_i, x_j)$ is independent of the time it would have taken j to be infected under a *different treatment* $x'_j \neq x_j$. We call this assumption a "cross-world" assumption because it makes explicit a probabilistic relationship between variables that cannot co-exist in the same realization of the process, namely $Z_i(w_j, \mathbf{x})$ or $W_i(x'_j)$.

Finally, let $T_i(W_j(x_j), \mathbf{x})$ be the potential outcome for the infection time of subject i, when j is infected at time $W_j(x_j)$ and the assigned treatments are $\mathbf{x} = (x_i, x_j)$. Following the decomposition (1) and by Assumptions 1–3, we can construct the potential infection time $T_i(W_j(x_j), \mathbf{x})$ as follows:

$$T_i(W_j(x_j), \mathbf{x}) = \begin{cases} W_i(x_i) & \text{if } W_i(x_i) < W_j(x_j) \\ W_j(x_j) + Z_i(W_j(x_j), \mathbf{x}) & \text{otherwise.} \end{cases} \quad (2)$$

The potential infection time with $W_j = w_j$ fixed is denoted as $T_i(w_j, \mathbf{x})$.

For convenience, define the binary potential infection outcome evaluated at time t, $Y_i(t; w_j, \mathbf{x}) = \mathbb{1}\{T_i(w_j, \mathbf{x}) < t\}$. We refer to the potential infection time $T_i(w_j, \mathbf{x})$ and infection outcome $Y_i(t; w_j, \mathbf{x})$ as *exposure-controlled* potential outcomes because they hold the

partner's infection time $W_j = w_j$ constant, thereby controlling the exposure to infection experienced by i. Similarly, we define $Y_i\left(t; W_j\left(x'_j\right), \mathbf{x}\right) = 1\left\{T_i\left(W_j\left(x'_j\right), \mathbf{x}\right) < t\right\}$, and refer to $T_i\left(W_j\left(x'_j\right), \mathbf{x}\right)$ and $Y_i\left(t; W_j\left(x'_j\right), \mathbf{x}\right)$ as *natural* potential outcomes because they do not control the exact infection time w_j of the partner, and instead rely on the natural distribution of W_j under the treatment x_j.

The potential infection time decomposition (2) formalizes intuition about the structure of interference under contagion: there can be no interference without prior infection. When neither i nor j is infected, the time to infection of i is solely a function of the treatment x_i, and there is no interference within the partnership. This is because the treatment x_j of j can only affect i *after* j becomes infected. When j is the first to be infected, the remaining time to infection of i is now a function of both x_i and x_j, because j has now gained the ability to transmit to i. This apparent complexity simplifies identification of causal effects, as we show below.

3 Causal estimands

Contrasts of potential infection outcomes under different treatments $\mathbf{x}$ and infection times w_j can yield epidemiologically meaningful causal estimands. In this paper, we express causal contrasts as differences of average potential infection outcomes. Effect measures on the hazard ratio, risk ratio, or odds ratio scales may be defined similarly [e.g. 24, 45].

First, the contagion effect captures the change in infection risk in one individual due to a change in the infection time of their partner.

Definition 1 (Contagion effect).

For $w_j \neq w'_j$ *and treatment* $\mathbf{x} = (x_i, x_j)$*, the controlled contagion effect is*
$\mathrm{CE}\left(t, w_j, w'_j, \mathbf{x}\right) = \mathbb{E}\left[Y_i\left(t; w_j, \mathbf{x}\right) - Y_i\left(t; w'_j, \mathbf{x}\right)\right]$ *and the natural contagion effect is*
$\mathrm{CE}(t, \mathbf{x}) = \mathbb{E}\left[Y_i\left(t; W_j(0), \mathbf{x}\right) - Y_i\left(t; W_j(1), \mathbf{x}\right)\right]$.

We say that the infection outcome (absent treatment) is "positively contagious" if for all infection times $w_j < w'_j$ with $w_j < t$, the controlled contagion effect under no treatment is $\mathrm{CE}\left(t, w_j, w'_j, \mathbf{0}\right) > 0$. In this way, we interpret contagion, or outcome transmissibility, as a causal phenomenon that need not depend on treatments: under positive contagion, earlier infection of one's partner *causes* one to become infected earlier, on average. On the other hand, the natural contagion effect $\mathrm{CE}(t, \mathbf{x})$ incorporates features of the treatment effect: it replaces fixed values of w_j and w'_j with their counterfactual distributions $W_j(0)$ and $W_j(1)$ when j is treated versus untreated, similar to the effect proposed by VanderWeele et al. [66] for an asymmetric partnership. The natural contagion effect is a "cross-world" estimand because it integrates the average potential infection outcome $\mathbb{E}\left[Y_i\left(t; w_j, \mathbf{x} = \left(x_i, x_j\right)\right)\right]$ with respect to the distribution of $W_j\left(x'_j\right)$ under a treatment $X_j = x'_j$ that cannot arise in the same realization as $X_j = x_j$. Figure 1 can be reinterpreted in light of Definition 1: positive

contagion means that earlier infection of j causes i to become infected earlier, compared to the infection time of i that would have occurred, had W_j happened later.

The susceptibility effect is of interest in vaccine trials because it summarizes the clinical effect of an intervention on the individual who receives it, holding the treatment status and infection time of their partner constant [18, 23, 26]. The susceptibility effect is sometimes called the "per-exposure effect" because it holds the distribution of exposure to infectiousness constant [45].

Definition 2 (Susceptibility effect).

For $w_j > 0$ *and* $X_j = x_j$*, the controlled susceptibility effect is*
$\mathrm{SE}(t, w_j, x_j) = \mathbb{E}\left[Y_i(t; w_j, x_i = 1, x_j) - Y_i(t; w_j, x_i = 0, x_j)\right]$ and the natural susceptibility effect is
$\mathrm{SE}(t, x_j) = \mathbb{E}\left[Y_i(t; W_j(x_j), x_i = 1, x_j) - Y_i(t; W_j(x_j), x_i = 0, x_j)\right]$.

If the controlled susceptibility effect is negative for every w_j and x_j, this means that the treatment is beneficial to the individual who receives it. Note that the natural susceptibility effect is not a cross-world estimand: it averages potential infection outcomes with respect to the distribution of $W_j(x_j)$, where x_j is the treatment under which the infection outcome of i is realized.

The infectiousness effect summarizes the effect of changing the treatment to j on the infection risk of i, while holding the treatment to i and the infection time of j unchanged.

Definition 3 (Infectiousness effect).

For $w_j > 0$ *and* $X_i = x_i$*, the controlled infectiousness effect is*
$\mathrm{IE}(t, w_j, x_i) = \mathbb{E}\left[Y_i(t; w_j, x_i, x_j = 1) - Y_i(t; w_j, x_i, x_j = 0)\right]$ *and the natural infectiousness effect is*
$\mathrm{IE}(t, x_i) = \mathbb{E}\left[Y_i(t; W_j(0), x_i, x_j = 1) - Y_i(t; W_j(0), x_i, x_j = 0)\right]$.

The natural infectiousness effect is a cross-world estimand because the first term in the contrast specifies that the infection time of j is realized under $x_j = 0$, but the infectiousness of j subsequently is realized under $x_j = 1$. Several authors have described the natural infectiousness effect as unidentified even under randomization when only binary infection outcomes are recorded at follow-up [10, 10–12, 20, 64, 66].

4 Identification of potential infection outcomes

We wish to non-parametrically identify the average potential infection outcome $\mathbb{E}\left[Y_i(t; w_j, \mathbf{x})\right]$ using observations of pairwise infection times, treatments, and covariates (T_i, T_j, X_i, X_j, $\mathbf{L}_i$, $\mathbf{L}_j$). A preliminary result identifies the distribution of $W_i(x_i)$ in Lemma 1 using information about infection times. The proof is given in the Appendix.

Lemma 1.

Suppose Assumptions 1, 3–5 *hold. Then the distribution function of* $W_i(x_i)$ *given* $\mathbf{L}_i = \mathbf{l}_i$ *is identified by*

$$F_i(w \mid x_i, \mathbf{l}_i) = 1 - \exp\left[-\int_0^w \frac{p\left(T_i = u, T_j > u \mid \mathbf{X} = \left(x_i, x_j\right), \mathbf{L} = \left(\mathbf{l}_i, \mathbf{l}_j\right)\right)}{\Pr\left(T_i > u, T_j > u \mid \mathbf{X} = \left(x_i, x_j\right), \mathbf{l} = \left(\mathbf{l}_i, \mathbf{l}_j\right)\right)} du\right]$$

for any fixed values of x_j, *and* $\mathbf{l}_j$, *where* $p(T_i = u, T_j > u|\mathbf{X} = (x_i, x_j), \mathbf{L} = (\mathbf{l}_i, \mathbf{l}_j))$ *is the joint probability density of* T_i *and survivor function of* T_j.

Lemma 1 is a standard distributional identification result in competing risks [2]. Here, W_i and W_j are competing event times within the same partnership. The distribution of W_i or W_j is identified utilizing both waiting times in the partnerships, even when the waiting times are censored due to lost to follow-up or administrative censoring for some partnerships. The identified distribution function $F_i(w|x_i, \mathbf{l}_i)$ is a function of x_i and $\mathbf{l}_i$ only, and is invariant to values of x_j and $\mathbf{l}_j$. However, in order to identify this function in the presence of the competing event W_j, particular values of x_j and $\mathbf{l}_j$ must be held constant.

The main result shows that average exposure-controlled potential infection outcomes given $\mathbf{L} = \mathbf{l}$ are identified. Proofs are given in the Appendix.

Theorem 1 (Identification of the average exposure-controlled potential infection outcome).

Suppose Assumptions 1–5 *hold and* $\mathbf{x} = (x_1, x_2)$. *For fixed values of* w_j *and* t, *if* $w_j < t$,

$$\mathbb{E}\left[Y_i\left(t; w_j, \mathbf{x}\right) \mid \mathbf{L} = \mathbf{l}\right] = F_i\left(w_j \mid x_i, \mathbf{l}_i\right) + \left(1 - F_i\left(w_j \mid x_i, \mathbf{l}_i\right)\right)\mathbb{E}\left[Y_i(t) \mid T_i \geq w_j, T_j = w_j, \mathbf{X} = \mathbf{X}, \mathbf{L} = \mathbf{l}\right] \tag{3}$$

otherwise, if $t \leq w_j$, $\mathbb{E}\left[Y_i\left(t; w_j, x\right) \mid \mathbf{L} = \mathbf{l}\right] = F_i\left(t \mid x_i, \mathbf{l}_i\right)$.

In Theorem 1, $F_i(w_j|x_i, \mathbf{l}_i)$ is identified by Lemma 1 using all infection times (including censored infection times), and $\mathbb{E}\left[Y_i(t) \mid T_i \geq w_j, T_j = w_j, \mathbf{X} = x, \mathbf{L} = \mathbf{l}\right]$ is estimated by the average outcome $Y_i(t)$ among observations when $T_j = w_j$, $T_i > T_j$ under $\mathbf{X} = \mathbf{x}$ and $\mathbf{L} = \mathbf{l}$.

The structure of (3) shows that the average exposure-controlled potential infection outcome is identified by two types of observable events: when i is infected before their partner, and when i is infected after their partner. In contrast to most work studying causal effect of vaccine using binary infection outcomes by the end of observation, the causal identification in (3) is built on observation of infection time, which provides sufficient control for exposure to infection. Figure 2 shows a causal diagram [48] that captures the causal structure among the variables in the system outlined by Assumptions 1–5. This causal diagram does not necessarily represent a causal non-parametric structural equation model (NPSEM). The approach proposed in this paper is not contingent on having a well-defined joint (probabilistic) density of the counterfactuals under every possible intervention, whereas, Shpitser et al. [57] build on NPSEMs that are represented using such causal diagrams.

If we do not fix the infection time $W_j = w_j$, and instead allow it to take its "natural" value under a particular treatment to j, we obtain the marginal average potential infection outcome when $\mathbf{L} = \mathbf{l}$ as follows.

Corollary 1 (Identification of average natural/exposure-marginalized potential infection outcome).

Suppose Assumptions 1–5 *hold. Then for* $\mathbf{x} = (x_i, x_j)$, $\mathbb{E}\left[Y_i\left(t; W_j\left(x_j\right), \mathbf{X}\right) \mid \mathbf{L} = \mathbf{l}\right] = \mathbb{E}\left[Y_i(t) \mid \mathbf{X} = x, \mathbf{L} = \mathbf{l}\right]$. *If in addition* $x'_j \neq x_j$ *and* Assumption 6 *holds,*

$$\mathbb{E}\left[Y_i\left(t; W_j\left(x'_j\right), \mathbf{x}\right) \mid \mathbf{L} = \mathbf{l}\right] = \int_0^t \mathbb{E}\left[Y_i\left(t; w_j, \mathbf{x}\right) \mid \mathbf{L} = \mathbf{l}\right] dF_j\left(w_j \mid x'_j, \mathbf{l}_j\right).$$

where $F_j\left(w'_j \mid x_j, \mathbf{l}_j\right)$ is given by Lemma 1 and $\mathbb{E}\left[Y_i\left(t; w_j, \mathrm{x}\right) \mid \mathbf{L} = \mathbf{l}\right]$ by Theorem 1.

Definition 3 and Corollary 1 together show why the natural infectiousness effect is not identified even under randomization when only binary infection outcomes are recorded at follow-up [10–12, 20, 64, 66]. The correct marginalization over infection times $W_j\left(x'_j\right)$ cannot be computed unless the distribution of $W_j\left(x'_j\right)$ is identified as in Lemma 1. The controlled and natural infectiousness effects are similar to those proposed by Chiba and Taguri [12], but here the marginalization is over the infection time of *j*, not their binary infection outcome.

Finally, by standardization of the potential infection outcome across the distribution of covariates **L**, we can identify the average potential infection outcome. Let $G(\mathbf{l})$ be the distribution function of the joint covariate vector $\mathbf{L} = \mathbf{l}$ in the population of partnerships. Then

$$\mathbb{E}\left[Y_i\left(t; w_j, \mathbf{x}\right)\right] = \int \mathbb{E}\left[Y_i\left(t; w_j, \mathbf{x}\right) \mid \mathbf{L} = \mathbf{l}\right] dG(\mathbf{l}) \tag{4}$$

and

$$\mathbb{E}\left[Y_i\left(t; W_j\left(x'_j\right), \mathbf{x}\right)\right] = \int \mathbb{E}\left[Y_i\left(t; W_j\left(x'_j\right), \mathbf{x}\right) \mid \mathbf{L} = \mathbf{l}\right] dG(\mathbf{l}) \tag{5}$$

where $\mathbb{E}\left[Y_i\left(t; w_j, \mathrm{x}\right) \mid \mathbf{L} = \mathbf{l}\right]$ and $\mathbb{E}\left[Y_i\left(t; W_j\left(x'_j\right), \mathrm{x}\right) \mid \mathbf{L} = \mathbf{l}\right]$ are given by Theorem 1 and Corollary 1 respectively. Because this paper is focused on nonparametric identification, we leave discussion of non-parametric statistical estimation of both controlled and natural causal estimands to the Appendix.

5 Comparison to other infectious disease intervention effects

Statisticians and epidemiologists have proposed a wide variety of estimands summarizing the effect of interventions for contagious outcomes, often in the two-person partnership setting. In this section, we evaluate the meaning of alternative definitions of vaccine effect estimands in the context of the causal effects defined above. We take the controlled contagion, susceptibility, and infectiousness effects defined above as fundamental characteristics of the disease transmission process and intervention under study. Whenever

possible, we characterize the sign, or direction, of alternative effects, as a function of these primitives. In some cases, where the relationship is complex, we evaluate the alternative estimands under a null hypothesis, for example when the controlled susceptibility or infectiousness effect is zero, so that explicit results can be analytically proven. For simplicity, we omit the role of covariates **L** in the comparison of estimands.

The "attack rate" of an infectious disease is defined for individuals with treatment x as $\mathrm{AR}_x(t) = \mathbb{E}\left[Y_i(t) \mid X_i = x\right]$. The ratio of attack rates, sometimes called "relative cumulative incidence", is a traditional measurement for the vaccine effect on susceptibility [12, 16, 17, 19, 24–27, 29, 30, 38, 47], defined as $\mathrm{VE}_{\mathrm{AR}}(t) = 1 - \mathrm{AR}_1(t)/\mathrm{AR}_0(t)$. A related estimand, called the "direct effect", is a contrast on the difference scale, $\mathrm{DE}(t) = \mathrm{AR}_1(t) - \mathrm{AR}_0(t)$ when treatment is randomized within groups [31]. In the symmetric partnership setting, attack rates $\mathrm{AR}_x(t)$ that condition only on the treatment to i implicitly marginalize over treatment to j.

Theorem 2.

Suppose $\mathrm{SE}(t, w_j, x_j) = 0$ *and* $\mathrm{IE}(t, w_j, x_j) < 0$ *for all* x_j *and* $w_j > 0$. *If* $\mathbf{X} = (X_i, X_j)$ *is positively dependent, then* $\mathrm{DE}(t) < 0$ *and* $\mathrm{VE}_{\mathrm{AR}}(t) > 0$*; if* $\mathbf{X}$ *is negatively dependent then* $\mathrm{DE}(t) > 0$ *and* $\mathrm{VE}_{\mathrm{AR}}(t) < 0$*; and if* $X_i \perp\!\!\!\perp X_j$ *then* $\mathrm{DE}(t) = \mathrm{VE}_{\mathrm{AR}}(t) = 0$. *If there is no treatment effect whatsoever,* $\mathrm{SE}(t, w_j, x) = \mathrm{IE}(t, w_j, x) = 0$ *for all* x *and* $w_j > 0$*, then* $\mathrm{DE}(t) = \mathrm{VE}_{\mathrm{AR}}(t) = 0$ *for any joint distribution of* $\mathbf{X}$.

In other words, $\mathrm{VE}_{\mathrm{AR}}(t)$ and $\mathrm{DE}(t)$ may or may not recover the sign, or direction, of the susceptibility effect, depending on the susceptibility and infectiousness effects, and the joint distribution of **X** within clusters. Morozova et al. [40] and Eck et al. [15] proved similar results in a parametric setting under Bernoulli, block, and cluster randomization for the joint treatment **X** in clusters or partnerships. Longini et al. [38], Halloran et al. [24], Halloran et al. [25], Halloran et al. [26] and Rhodes et al. [51] warned that $\mathrm{VE}_{\mathrm{AR}}(t)$ may be a biased approximation to the susceptibility effect due to differential exposure to infection between treated and untreated individuals in clusters. We show simulation examples that result in biased $DE(t)$ under block randomization in Table 1 and Figure 4(d) below.

Related definitions of the attack rate condition on the treatments to both individuals in the partnership. The attack rate among individuals with treatment x whose partner has treatment x' is $\mathrm{AR}_{x,x'}(t) = \mathbb{E}\left[Y_i(t) \mid X_i = x, X_j = x'\right]$. The indirect effect is defined as $\mathrm{IDE}(t) = \mathrm{AR}_{01}(t) - \mathrm{AR}_{00}(t)$ [12, 31], and is equivalent to the difference of the natural infectiousness and contagion effects defined above:

$$\begin{aligned}\mathrm{IDE}(t) &= \mathbb{E}\left[Y_i\left(t; W_j(1), (0,1)\right) - Y_i\left(t; W_j(0), \mathbf{0}\right)\right] \\ &= \mathbb{E}\left[Y_i\left(t; W_j(1), (0,1)\right) - Y_i\left(t; W_j(0), (0,1)\right)\right] + \mathbb{E}\left[Y_i\left(t; W_j(0), (0,1)\right) - Y_i\left(t; W_j(0), \mathbf{0}\right)\right] \\ &= -\mathrm{CE}(t, (0,1)) + \mathrm{IE}(t, 0)\,.\end{aligned}$$

The secondary attack rate is the proportion in a cluster infected after being exposed to an earlier infected individual, formally defined as

$\mathrm{SAR}_{x',x}(t) = \mathbb{E}\left[Y_i(t) \mid T_j < t, T_i > T_j, X_i = x, X_j = x'\right]$. The SAR is the average infection status of i when j is infected during the study before i, under treatments x and x' to i and j respectively. Based on the potential pitfalls of SAR, researchers proposed $\mathrm{VE}_{\mathrm{I}}^{\mathrm{net}}(t) = 1 - \mathrm{SAR}_{10}(t)/\mathrm{SAR}_{00}(t)$ as "secondary attack rate for infectiousness" Halloran and Hudgens [20], Halloran et al. [24, 26, 27, 28, 29], Orenstein et al. [47]. We analyze $\mathrm{VE}_{\mathrm{I}}^{\mathrm{net}}(t)$ under the null hypothesis of no infectiousness effect, and show that when the infection is contagious and there is a susceptibility effect, $\mathrm{VE}_{\mathrm{I}}^{\mathrm{net}}(t)$ may nevertheless be nonzero. Let $h_0(u|0)$ be the hazard of the potential infection time $W_i(0)$, and let $h_0(u|1)$ be the hazard of $W_i(1)$.

Theorem 3.

Suppose $\mathrm{IE}(t, w_j, 0) = 0$, $\mathrm{CE}\left(t, w_j, w'_j, \mathbf{0}\right) > 0$ *for all* $0 < w_j < w'_j$, *and* $h_0(u|1) = \varepsilon h_0(u|0)$ *with* $\varepsilon \in [0,1)$, *then* $\mathrm{VE}_{\mathrm{I}}^{\mathrm{net}}(t) > 0$. *If* $\mathrm{SE}(t, w_j, x_j) = 0$ *for all* w_j *and* x_j, *then* $\mathrm{VE}_{\mathrm{I}}^{\mathrm{net}}(t)$ *preserves the same sign as* $\mathrm{IE}(t, w_j, 0)$. *Suppose* $\mathrm{CE}\left(t, w_j, w'_j, \mathbf{0}\right) = 0$ *for all* $0 < w_j < w'_j$ *and* $h_0(u|1) = \varepsilon h_0(u|0)$ *with* $\varepsilon \in [0,1)$, *then* $\mathrm{VE}_{\mathrm{I}}^{\mathrm{net}}(t) > 0$.

In other words, when the true infectiousness effect is null, the infection outcome is positively contagious, and the vaccine has a favorable susceptibility effect prior to the first infection, $\mathrm{VE}_{\mathrm{I}}^{\mathrm{net}}(t)$ can nevertheless be nonzero. In a more extreme case, when the true contagion effect is null, the disease is not transmissible so that the true infectiousness is null; if the vaccine has a favorable susceptibility effect prior to the first infection, then $\mathrm{VE}_{\mathrm{I}}^{\mathrm{net}}(t)$ is still nonzero. Simulation examples show biased $\mathrm{VE}_{\mathrm{I}}^{\mathrm{net}}(t)$ under a null contagion effect in Tables 1 and 2 below.

A simple explanation shows why $\mathrm{VE}_{\mathrm{I}}^{\mathrm{net}}$ can behave in unexpected ways: it is not solely a function of the infectiousness effect. Instead, $\mathrm{VE}_{\mathrm{I}}^{\mathrm{net}}(t)$ also incorporates reduced exposure to infection from delaying the infection of partner j due to vaccination, which in fact is the susceptibility effect on the partner j *before the first infection occurs*. Therefore, when the true susceptibility effect is null, $\mathrm{VE}_{\mathrm{I}}^{\mathrm{net}}(t)$ is only a function of the infectiousness effect and thus recovers the correct sign of infectiousness effect. From a sightly different perspective, several authors have also pointed out that $\mathrm{VE}_{\mathrm{I}}^{\mathrm{net}}(t)$ may suffer from selection bias because it conditions on post-randomization variables – the infection status of both partners [20, 24–26, 51]. Specifically, $\mathrm{VE}_{\mathrm{I}}^{\mathrm{net}}(t)$ relies on the eventual infection outcome of partner j, rather than the infection time of partner j. Halloran and Hudgens [20] use tools from principal stratification to derive bounds for the infectiousness effect to correct this selection bias, and propose a bound estimator $\mathrm{CVE}_{\mathrm{I}}^{c}(t)$ for $\mathrm{VE}_{\mathrm{I}}^{\mathrm{net}}(t)$ under Bernoulli randomization. We analyze these bounds by simulation below.

Several authors have recognized that simple comparison of outcomes in treated versus untreated individuals may not suffice to identify meaningful causal effects for infectious disease interventions, even under randomization. For example, VanderWeele [63], VanderWeele et al. [66], Ogburn and VanderWeele [42], and Ogburn and VanderWeele [43] apply tools from mediation analysis to a simplified partnership model to identify contagion and infectiousness effects similar to those we have defined above. This "asymmetric partnership" model focuses on pairs of individuals i and j when i is restricted to be home-bound, unvaccinated, and may only be infected by their (possibly vaccinated) partner j. Partner j is randomized to receive treatment or placebo, and may be infected by a source of infection outside the partnership. In other words, the relative role of the two subjects cannot be swapped. For example, in a HIV trial of zidovudine, the study units are mother-child pairs, and only mothers are vaccinated and may transmit HIV to the children, not vice versa [39]. This is different from the symmetric partnership setting we considered, when both i and j can be treated and infected by the outside or each other.

To represent this structural assumption in the framework outlined here, we force the infection time of the home-bound partner, in the absence of infection in their partner, to be infinite. To this end, let hazard of $W_i(0)$ be $h_0^i(t \mid 0) = 0$, so that infection of i from an external source can never occur. These authors define the infectiousness effect as $\mathrm{VE}_\mathrm{I}(t) = \mathbb{E}\left[Y_i\left(t; Y_j(1), (0,1)\right)\right] - \mathbb{E}\left[Y_i\left(t; Y_j(1), (0,0)\right)\right]$, which contrasts the infection outcomes of i when j is treated versus untreated, with j's infection status $Y_j(x_j)$ set to the value it would take if j were treated.

Theorem 4.

Suppose $h_0^i(t \mid 0) = 0$ *for all* $t > 0$. *Then* $\mathrm{VE}_\mathrm{I}(t) = \mathrm{IE}(t, 0)$.

In other words, under the asymmetric setting where i is unvaccinated and cannot be infected from outside the partnership, $\mathrm{VE}_\mathrm{I}(t)$ is equivalent to the natural infectiousness effect in Definition 3.

A contagion effect is defined by VanderWeele et al. [66] as

$$\mathrm{VE}_\mathbf{C}(t) = \mathbb{E}\left[Y_i\left(t; Y_j(1), (0,0)\right) - Y_i\left(t; Y_j(0), (0,0)\right)\right],$$

contrasting the infection outcome of i when the infection status of j is set to the value it would obtain if j were treated versus untreated. Note that this quantity reverses the difference in the natural contagion effect in Definition 1, as $\mathrm{VE}_\mathbf{C}(t) = -CE(t, \mathbf{x})$. We provide sufficient conditions for the controlled contagion effect $\mathrm{CE}(t, u, u', \mathbf{0})$ and $\mathrm{VE}_\mathbf{C}(t)$ (or equivalently, $-CE(t, \mathbf{x})$) to behave similarly, that is, to have opposite sign.

Theorem 5.

Suppose $h_0^i(t \mid 0) = 0$ *for all* $t > 0$ *and* $\mathrm{SE}(t, w_j, 0) > 0$. *Then* $\mathrm{VE}_{\mathbf{C}}(t)$ *has opposite sign as* $\mathrm{CE}\left(t, w_j, w'_j, \mathbf{0}\right)$ *for* $0 < wj < w'_j$. *Suppose* $h_0^i(t \mid 0) = 0$, $\mathrm{SE}(t, w_j, 0) = 0$ *and* $\mathrm{CE}\left(t, w_j, w'_j, \mathbf{0}\right) > 0$, *then* $\mathrm{VE}_{\mathbf{C}}(t) = 0$.

In other words, in the asymmetric partnership setting, $-\mathrm{VE}_{\mathbf{C}}(t)$ recovers the sign of the true contagion effect, when the vaccine has a favorable susceptibility effect *prior to the first infection*. However, if the true susceptibility effect is null, $\mathrm{VE}_{\mathbf{C}}(t) = 0$ regardless of the true contagion effect.

6 Application: a hypothetical vaccine trial

We simulate observational and randomized trials of a hypothetical HIV vaccine in a large population of sexual partnerships [25]. We assume individuals are not infected at baseline, but that either individual may become infected from outside the partnership, and transmission within partnerships may occur. To parameterize the infection transmission process, we specify hazard models for the infection times $W_i(x_i)$ and $Z_i(w_j, \mathbf{x})$. This approach has been employed in extensive prior work on statistical models for time-to-infection data [25, 32–35, 51]. For a time $t > 0$, Let the hazard of $W_i(x_i)$ given covariates $\mathbf{L}_i = l_i$ be given by

$$\lambda_i^W\left(t; x_i, l_i\right) = \alpha(t) e^{\beta_0 x_i + \theta'_0 \mathbf{l}_i}. \tag{6}$$

In words, the hazard of infection in an individual whose partner is not infected, is given by a Cox model with baseline hazard $\alpha(t)$. Following infection of j at time $W_j = w_j$, the remaining potential infection time $Z_i(w_j, \mathbf{x})$ given $\mathbf{L} = \mathbf{l} = (l_i, l_j)$ has hazard

$$\lambda_i^Z\left(t; w_j, \mathrm{x}, \mathbf{l}\right) = \lambda_i^W\left(t; x_i\right) + \gamma\left(t - w_j\right) e^{\beta_1 x_i + \sigma x_j + \theta'_1 \mathbf{l}_j + \theta'_2 \mathbf{l}_i} \tag{7}$$

for $t > w_j$. The coefficients β_0 and β_1 represent the change in infection risk due to vaccination of i, and σ represents the change in transmission risk due to vaccination in j when j is infected. Covariate effects are represented by θ_0, θ_1, and θ_2, and $\alpha(t)$ and $\gamma(t - w_j)$ are baseline transmission hazards for the external and internal forces of infection respectively. This specification implies that the external force of infection and transmissibility are *competing risks* for infection of i [37, 38, 50]. That is, a susceptible individual can be infected by a source of infectiousness outside their partnership, or from an infected partner. We consider three specifications of the baseline transmission hazards for the external and internal forces of infection: (i) both are time-invariant as in (8)

$$\begin{aligned} \alpha(t) &= \alpha \\ \gamma(t) &= \gamma \end{aligned} \tag{8}$$

, (ii) the external baseline hazard varies seasonally and the internal baseline hazard decays over time as in (9)

$$\begin{aligned} \alpha(t) &= a(1+\sin(2\pi t+\phi)) \\ \gamma\left(t-w_j\right) &= b\exp\left[-\omega\left(t-w_j\right)\right] \end{aligned} \tag{9}$$

, to (iii) when the external baseline hazard varies over seasons and the internal baseline hazard increases first then decreases over time as in (10).

$$\begin{aligned} \alpha(t) &= a(1+\sin(2\pi t+\phi)) \\ \gamma\left(t-w_j\right) &= b\frac{1}{\Gamma(k)\theta^k}\left(t-t_j\right)^{k-1}e^{-\frac{t-t_j}{\theta}} \end{aligned} \tag{10}$$

When the baseline hazards $\alpha(t)$ and $\gamma(t-w_j)$ are time-invariant as specified in (8), the model reduces to a Markov susceptible-infective process with an external force of infection [e.g 15, 40]. For any functional forms of the baseline hazards $\alpha(t)$ and $\gamma(t-w_j)$, the hazard specifications (6) and (7) imply distributions for $W_i(x_i)$ and $Z_i(w_j, \mathbf{x})$, and hence $T_i(w_j, \mathbf{x})$, that obey the required identification Assumptions 1–6.

Subjects in partnerships are endowed with individual characteristics $\mathbf{L} = (\mathbf{L}_i, \mathbf{L}_j)$ that may be correlated. In the randomized trial simulation, the vaccine is randomized in accordance with a specified distribution – Bernoulli, block, or cluster randomization – without regard to these traits. Under each randomization design, the marginal treatment probability $\Pr(X_i = x_i)$ is 1/2. For Bernoulli randomization, $\Pr(\mathbf{X} = \mathbf{x}) = 1/4$, for block randomization, $\Pr(x = x) = 1\left\{\sum_i x_i = 1\right\}/2$, and for cluster randomization, $\Pr(\mathbf{X} = (1, 1)) = 1/2$ and $\Pr(\mathbf{X} = (0, 0)) = 1/2$. In the observational study simulation, we consider a univariate individual covariate for illustration, and the traits $\mathbf{L} = (L_i, L_j)$ together determine the joint distribution of vaccine in the partnership as

$$\Pr\left(X_i = 1 \mid L_i = l_i\right) = \frac{1}{1+e^{-l_i}}$$

where

$$\begin{pmatrix} L_i \\ L_j \end{pmatrix} \sim \text{Normal}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, v\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$$

with $v > 0$. Non-parametric estimation of both controlled and natural causal estimands is described in detail in the Appendix.

Figure 3 illustrates controlled infection outcomes $\mathbb{E}\left[Y_i\left(t; w_j, x\right)\right]$ over time for different choices of w_j and $\mathbf{x}$ under the time-invariant hazard scenario, estimated non-parametrically with sufficiently large numbers of pairs ($N = 100,000$) so as to represent their true values in the simulation. Estimated controlled infection outcomes area aligned with their true values

in Figure 3. Contrasts of these potential infection outcomes give the controlled contagion, susceptibility and infectiousness effects, shown in the lower-right corner of Figure 3.

Tables 1 and 2 show the true values of the natural contagion, susceptibility and infectiousness effects, and compare these values to the true values of alternative estimands proposed by other authors, including the direct effect DE(t), the indirect effect IDE(t), the secondary attack rate infectiousness effect $\mathrm{VE}_{\mathrm{I}}^{\mathrm{net}}(t)$, and $\mathrm{CVE}_{\mathrm{I}}^{c}(t)$ bounds introduced by [20]. All natural or marginal estimands are evaluated at time $t = 2$ years under each design and under both time-invariant and time-varying baseline hazards. Estimands that are not identified under a given design are not evaluated. In Table 1, when the true infectiousness effect much stronger than the true susceptibility effect, the direct effect DE(t) is positive (0.06 and 0.08) under block randomization when the disease is contagious, even though the true susceptibility effect is negative, or beneficial [see, e.g. 15, 40]. Table 2 shows another simulation setting where DE(t) achieves the same sign as the susceptibility effect when the true infectiousness effect is on the same scale of the true susceptibility effect. In the three scenarios without contagion, the disease is not contagious and infection outcomes are realized independently. Therefore, all "indirect" and "infectiousness" effects should be null. However, $\mathrm{VE}_{\mathrm{I}}^{\mathrm{net}}(t)$ is negative (−0.01 and −0.02 in both Table 1 and 2), conflicting with the fact that the disease is not transmissible (as proved in Theorem 3). The identification interval $\mathrm{CVE}_{\mathrm{I}}^{c}(t)$ has nonzero width, but covers zero.

Figure 4 compares different types of natural susceptibility and infectiousness effects over time, when both effects are beneficial (negative). In the bottom-right panel of Figure 4, we show that DE(t) under block randomization can suffer from directional bias.

7 Discussion

We have described a nonparametric framework for identifying causal intervention effects under contagion in general two-person partnerships. The estimands and identification results generalize those given in prior work [20, 43, 64, 66], and establish that point identification of clinically meaningful causal estimands under contagion is possible even when relationships are symmetric and either individual can be treated. We take a nonparametric approach that does not ascribe infections to particular sources. Instead, the approach focuses on the effect of changing treatments or exposure to infections on the expectations of potential outcomes without information about "who-infected-whom." We have made no assumptions about the functional form of infection risks (beyond the independencies and exclusion restrictions implied by Assumptions 1–6), how the risk of infection to a susceptible individual changes when their partner becomes infected, or how the vaccine changes susceptibility or infectiousness over time. The framework respects the logic of infectious disease transmission: if the outcome is not transmissible, the contagion and infectiousness effects are zero.

By studying the role of a partner's infection time in the identification of controlled causal effects, we can identify causal estimands that are both more fundamental and more directly linked to the biological effect of a vaccine on infection risk than simple contrasts of infection

rates. Our results also show that while some crude contrasts can recover causal effects in restricted settings (e.g. the infectious effect $VE_I(t)$ in the asymmetric partnership setting) or under a particular randomization design (e.g. the direct effect $DE(t)$ under independent Bernoulli randomization), they may not deliver useful summaries of vaccine effects in more general situations. Finally, the framework developed in this paper may be useful in settings beyond infectious disease epidemiology, where symmetric mediated effects are of interest [e.g. 55, 58].

One important limitation of our identification approach is that the controlled estimands and cross-world natural estimands require observation of infection times, and not just binary infection indicators at a follow-up time t. In real-world vaccine trials, it may be unreasonable to require investigators to measure infection times T_i with precision, as is required by Lemma 1 and Theorem 1. Instead, cross-sectional infection assessment, follow-up surveys, or tests for biomarkers of prior infection are commonly used as the primary outcome. Corollary 1 shows exactly how controlled effects that rely on infection times relate to natural effects that do not. Attempts to disentangle individual effects from the mediating effects of treatment to partners using only binary infection outcomes may fail to recover useful controlled or marginal effects [see, e.g. VE_I^{net}, analyzed by 20]. One exception is the natural susceptibility effect, which can be estimated by binary outcomes under Bernoulli randomizations, as shown by Corollary 1.

Finally, while the symmetric partnership setting is useful for conceptualizing, defining, and identifying causal estimands, real-world vaccine trials usually happen in clusters of varying sizes. Adapting the setting outlined here to larger clusters results in rapid expansion of the number of potential outcomes, corresponding to every possible ordering of infections, necessitating simplifying structural assumptions to reduce the dimensionality of the problem. One promising avenue for dramatically reducing the number of potential outcomes without imposing a parametric structure was proposed by Kenah [33, 34]. The idea is that contagion works by competing risks, where hazards of infection from different sources are additive. This approach imposes no additional structure on the distribution of the initial time to infection, but assumes that new infected cluster members always add a competing risk of infection to the already existing risks of infection for susceptibles.

Acknowledgements

WWL was supported by NIH grant R01 AI085073 and by a Gillings Innovation Laboratory award from the UNC Gillings School of Global Public Health. FWC was supported by NIH grants DP2 OD022614, R01 AI112438, and R01 AI112970. We are grateful to Peter M. Aronow, Soheil Eshghi, Eben Kenah, Olga Morozova, Virginia E. Pitzer, and Li Zeng for helpful comments and discussion.

A: Proofs

Proof of Lemma 1.

Let $f_i(w|x_i, \mathbf{l}_i)$ be the density of $W_i(x_i)$ when $\mathbf{L}_i = \mathbf{l}_i$ and let $F_i(w|x_i, \mathbf{l}_i)$ be the corresponding cumulative distribution function. By Assumption 5, $0 < F_i(w|x_i, \mathbf{l}_i) < 1$ for all $w > 0$, x_i, and $\mathbf{l}_i$, so we can write

$$\frac{f_i(w \mid x_i, l_i)}{1 - F_i(w \mid x_i, l_i)} = -\frac{d}{dw}\log\left(1 - F_i(w \mid x_i, l_i)\right).$$

Then rearranging, we have

$$F_i(w \mid x_i, l_i) = 1 - \exp\left[-\int_0^w \frac{f_i(u \mid x_i, l_i)}{1 - F_i(u \mid x_i, l_i))} du\right]$$
$$= 1 - \exp\left[-\int_0^w \frac{f_i(u \mid x_i, \mathbf{l}_i)\left(1 - F_j(u \mid x_j, \mathbf{l}_j)\right)}{\left(1 - F_i(u \mid x_i, \mathbf{l}_i)\right)\left(1 - F_j(u \mid x_j, \mathbf{l}_j)\right)} du\right]$$

by Assumption 1

$$= 1 - \exp\left[-\int_0^w \frac{p\left(W_i(x_i) = u, W_j(x_j) > u \mid \mathbf{X} = (x_i, x_j), \mathbf{L} = (\mathbf{l}_i, \mathbf{l}_j)\right)}{\Pr\left(W_i(x_i) > u, W_j(x_j) > u \mid \mathbf{X} = (x_i, x_j), \mathbf{L} = (\mathbf{l}_i, \mathbf{l}_j)\right)} du\right]$$

by Assumption 3

$$= 1 - \exp\left[-\int_0^w \frac{p\left(W_i = u, W_j > u \mid \mathbf{X} = (x_i, x_j), \mathbf{L} = (\mathbf{l}_i, \mathbf{l}_j)\right)}{\Pr\left(W_i > u, W_j > u \mid \mathbf{X} = (x_i, x_j), \mathbf{L} = (\mathbf{l}_i, \mathbf{l}_j)\right)} du\right]$$

by Assumption 4

$$= 1 - \exp\left[-\int_0^w \frac{p\left(T_i = u, T_j > u \mid \mathbf{X} = (x_i, x_j), \mathbf{L} = (\mathbf{l}_i, \mathbf{l}_j)\right)}{\Pr\left(T_i > u, T_j > u \mid \mathbf{X} = (x_i, x_j), \mathbf{L} = (\mathbf{l}_i, \mathbf{l}_j)\right)} du\right],$$

where x_j is any fixed value of X_j and $\mathbf{l}_j$ is any fixed value of $\mathbf{L}_j$. □

Lemma 2.

Under Assumptions 1–3, $Y_i(t; w_j, \mathbf{x}) \perp\!\!\!\perp W_j(x_j) \mid \mathbf{L}$ *and* $Y_i(t; w_j, \mathbf{x}) \perp\!\!\!\perp \mathbf{X} \mid \mathbf{L}$.

Proof of Lemma 2.

Fix a value $w_j > 0$ and let $\mathbf{x} = (x_i, x_j)$. If $W_i(x_i) < w_j$, then $T_i(w_j, \mathbf{x}) = W_i(x_i)$ and by Assumption 1, $W_i(x_i) \perp\!\!\!\perp W_j(x_j) \mid \mathbf{L}$, so $T_i(w_j, \mathbf{x}) \perp\!\!\!\perp W_j(x_j) \mid \mathbf{L}$. If $W_i(x_i) > w_j$ then $T_i(w_j, \mathbf{x}) = w_j + Z_i(w_j, \mathbf{x})$ and by Assumption 2 $Z_i(w_j, \mathbf{x}) \perp\!\!\!\perp W_j \mid \mathbf{L}$, so $T_i(w_j, \mathbf{x}) \perp\!\!\!\perp W_j(x_j) \mid \mathbf{L}$. Therefore, since $Y_i(t; w_j, \mathbf{X}) = 1\left\{T_i(w_j, \mathbf{X}) < t\right\}$, it follows that $Y_i(t; w_j, \mathbf{x}) \perp\!\!\!\perp W_j(x_j) \mid \mathbf{L}$.

By the same reasoning, if $W_i(x_i) < w_j$, then $T_i(w_j, \mathbf{x}) = W_i(x_i)$ and by Assumption 3, $W_i(x_i) \perp\!\!\!\perp \mathbf{X} \mid \mathbf{L}$. If $W_i(x_i) > w_j$ then $T_i(w_j, \mathbf{x}) = w_j + Z_i(w_j, \mathbf{x})$ and by Assumption 3, $Z_i(w_j, \mathbf{x}) \perp\!\!\!\perp \mathbf{X} \mid \mathbf{L}$. Therefore, since $Y_i(t; w_j, \mathbf{X}) = 1\left\{T_i(w_j, \mathbf{X}) < t\right\}$, it follows that $Y_i(t; w_j, \mathbf{x}) \perp\!\!\!\perp \mathbf{X} \mid \mathbf{L}$. □

Lemma 3.

Under Assumptions 1–4, $\mathbb{E}\left[Y_i\left(t, w_j, x\right)\right] = \mathbb{E}\left[Y_i(t) \mid W_j = w_j, \mathbf{X} = \mathbf{X}\right]$.

Proof of Lemma 3.

Fix a value $w_j > 0$ and $\mathbf{x} = (x_i, x_j)$. If $W_i(x_i) \geq w_j$ then

$$\begin{aligned}
\mathbb{E}\left[Y_i\left(t, w_j, \mathrm{x}\right)\right] &= \Pr\left(T_i\left(w_j, \mathrm{x}\right) < t\right) \text{ by the definition of } Y_i\left(t, w_j, \mathrm{x}\right) \\
&= \Pr\left(w_j + Z_i\left(w_j, \mathrm{x}\right) < t\right) \text{ by the definition of } T_i\left(w_j, \mathrm{x}\right) \text{ and } W_i\left(x_i\right) \geq w_j \\
&= \Pr\left(Z_i\left(w_j, \mathrm{x}\right) < t - w_j\right) \\
&= \Pr\left(Z_i\left(w_j, \mathrm{x}\right) < t - w_j \mid W_j = w_j\right) \text{ by Assumption 2} \\
&= \Pr\left(Z_i\left(w_j, \mathrm{x}\right) < t - w_j \mid W_j = w_j, \mathbf{X} = \mathrm{x}\right) \text{ by Assumption 3} \\
&= \Pr\left(Z_i < t - w_j \mid W_j = w_j, \mathbf{X} = \mathrm{x}\right) \text{ by Assumption 4} \\
&= \Pr\left(Z_i < t - W_j \mid W_j = w_j, \mathbf{X} = \mathrm{x}\right) \\
&= \Pr\left(Z_i + W_j < t \mid W_j = w_j, \mathbf{X} = \mathrm{x}\right) \\
&= \Pr\left(T_i < t \mid W_j = w_j, \mathbf{X} = \mathrm{x}\right) \text{ by the definition of } T_i \\
&= \mathbb{E}\left[Y_i(t) \mid W_j = w_j, \mathbf{X} = \mathrm{x}\right] \text{ by the definition of } Y_i(t)
\end{aligned}$$

If $W_i(x_i) < w_j$ then

$$\begin{aligned}
\mathbb{E}\left[Y_i\left(t, w_j, \mathrm{x}\right)\right] &= \Pr\left(T_i\left(w_j, \mathrm{x}\right) < t\right) \text{ by the definition of } Y_i\left(t, w_j, \mathrm{x}\right) \\
&= \Pr\left(W_i\left(x_i\right) < t\right) \text{ by the definition of } T_i\left(w_j, \mathrm{x}\right) \text{ and } W_i\left(x_i\right) < w_j \\
&= \Pr\left(W_i\left(x_i\right) < t \mid X_i = x_i, X_j = x_j\right) \text{ by Assumption 3} \\
&= \Pr\left(W_i\left(x_i\right) < t \mid W_j = w_j, X_i = x_i, X_j = x_j\right) \text{ by Assumption 1} \\
&= \Pr\left(W_i < t \mid W_j = w_j, \mathbf{X} = \mathrm{x}\right) \text{ by Assumption 4} \\
&= \Pr\left(T_i < t \mid W_j = w_j, \mathbf{X} = \mathrm{x}\right) \text{ by the definition of } T_i \\
&= \mathbb{E}\left[Y_i(t) \mid W_j = w_j, \mathbf{X} = \mathrm{x}\right] \text{ by the definition of } Y_i(t)
\end{aligned}$$

□

Proof of Theorem 1.

The average potential infection outcome when $\mathbf{L} = \mathbf{l}$ is given by

$\mathbb{E}\left[Y_i\left(t; w_j, \mathrm{x}\right) \mid L = l\right] = \mathbb{E}\left[Y_i\left(t; w_j, x\right) \mid W_j = w_j, X = x, L = l\right]$ by Lemma 2

$$\begin{aligned}
&= \mathbb{E}\left[Y_i\left(t; w_j, \mathrm{x}\right) \mid W_i \leq w_j, W_j = w_j, \mathbf{X} = \mathrm{x}, \mathbf{L} = \mathbf{l}\right]\Pr\left(W_i \leq w_j \mid W_j = w_j, \mathbf{X} = \mathrm{x}, \mathbf{L} = \mathbf{l}\right)\Big) \\
&+ \mathbb{E}\left[Y_i\left(t; w_j, \mathrm{x}\right) \mid W_i > w_j, W_j = w_j, \mathbf{X} = \mathrm{x}, \mathbf{L} = \mathbf{l}\right]\Pr\left(W_i > w_j \mid W_j = w_j, \mathbf{X} = \mathrm{x}, \mathbf{L} = \mathbf{l}\right)\Big) \\
&= \Pr\left(T_i\left(w_j, \mathrm{x}\right) < t \mid W_i \leq w_j, W_j = w_j, \mathbf{X} = \mathrm{x}, \mathbf{L} = \mathbf{l}\right)\Pr\left(W_i \leq w_j \mid W_j = w_j, \mathbf{X} = \mathrm{x}, \mathbf{L} = \mathbf{l}\right)\Big) \\
&+ \mathbb{E}\left[Y_i\left(t; w_j, \mathrm{x}\right) \mid W_i > w_j, W_j = w_j, \mathbf{X} = \mathrm{x}, \mathbf{L} = \mathbf{l}\right]\Pr\left(W_i > w_j \mid W_j = w_j, \mathbf{X} = \mathrm{x}, \mathbf{L} = \mathbf{l}\right)\Big)
\end{aligned}$$

by the definition of $Y_i(t; w_j, \mathbf{x})$

$$= \Pr\left(W_i(x_i) < t \mid W_i \le w_j, W_j = w_j, \mathbf{X} = \mathrm{x}, \mathbf{L} = \mathbf{l}\right)\Pr\left(W_i \le w_j \mid W_j = w_j, \mathbf{X} = \mathrm{x}, \mathbf{L} = \mathbf{l}\right)\Big)$$
$$+\mathbb{E}\left[Y_i\left(t; w_j, \mathrm{x}\right) \mid W_i > w_j, W_j = w_j, \mathbf{X} = \mathrm{x}, \mathbf{L} = \mathbf{l}\right]\Pr\left(W_i > w_j \mid W_j = w_j, \mathbf{X} = \mathrm{x}, \mathbf{L} = \mathbf{l}\right)\Big)$$

by the definition of $T_i(w_j, \mathbf{x})$

$$= \Pr\left(W_i(x_i) < t \mid W_i \le w_j, X_i = x_i, \mathbf{L} = \mathbf{l}\right)\Pr\left(W_i \le w_j \mid X_i = x_i, \mathbf{L} = \mathbf{l}\right)\Big)$$
$$+\mathbb{E}\left[Y_i\left(t; w_j, \mathrm{x}\right) \mid W_i > w_j, W_j = w_j, \mathbf{X} = \mathrm{x}, \mathbf{L} = \mathbf{l}\right]\Pr\left(W_i > w_j \mid X_i = x_i, \mathbf{L} = \mathbf{l}\right)\Big)$$

by Assumption 1

$$= \Pr\left(W_i < t \mid W_i \le w_j, X_i = x_i, \mathbf{L} = l\right)\Pr\left(W_i \le w_j \mid X_i = x_i, \mathbf{L} = l\right)\Big)$$
$$+\mathbb{E}\left[Y_i(t) \mid W_i > w_j, W_j = w_j, \mathbf{X} = \mathrm{x}, \mathbf{L} = l\right]\Pr\left(W_i > w_j \mid X_i = x_i, \mathbf{L} = l\right)\Big)$$

by Assumption 4 and Lemma 3

$$= \Pr\left(W_i < t, W_j \le w_j \mid X_i = x_i, \mathbf{l} = l\right)$$
$$+\mathbb{E}\left[Y_i(t) \mid W_i > w_j, W_j = w_j, \mathbf{X} = \mathrm{x}, \mathbf{L} = \mathbf{l}\right]\Pr\left(W_i > w_j \mid X_i = x_i, \mathbf{L} = \mathbf{l}\right)\Big)$$

When $t \ge w_j$, then

$$\begin{aligned}\mathbb{E}\left[Y_i\left(t; w_j, \mathrm{x}\right) \mid \mathbf{L} = \mathbf{l}\right] &= \Pr\left(W_i < t, W_j \le w_j \mid X_i = x_i, \mathbf{L} = \mathbf{l}\right)\\ &\quad +\mathbb{E}\left[Y_i(t) \mid W_i > w_j, W_j = w_j, \mathbf{X} = \mathrm{x}, \mathbf{L} = \mathbf{l}\right]\Pr\left(W_i > w_j \mid X_i = x_i, \mathbf{L} = \mathbf{l}\right)\Big)\\ &= \Pr\left(W_i \le w_j \mid X_i = x_i, \mathbf{L} = \mathbf{l}\right)\Big)\\ &\quad +\mathbb{E}\left[Y_i(t) \mid W_i > w_j, W_j = w_j, \mathbf{X} = \mathrm{x}, \mathbf{L} = \mathbf{l}\right]\Pr\left(W_i > w_j \mid X_i = x_i, \mathbf{L} = \mathbf{l}\right)\Big)\\ &= F_i\left(w_j \mid x_i, \mathbf{l}_i\right) + \left(1 - F_i\left(w_j \mid x_i, \mathbf{l}_i\right)\right)\mathbb{E}\left[Y_i(t) \mid W_i > w_j, W_j = w_j, \mathbf{X} = \mathrm{x}, \mathbf{L} = \mathbf{l}\right].\end{aligned}$$

Likewise, when $t < w_j$, then

$$\begin{aligned}\mathbb{E}\left[Y_i\left(t; w_j, \mathrm{x}\right) \mid \mathbf{L} = \mathbf{l}\right] &= \Pr\left(W_i < t, W_j \le w_j \mid X_i = x_i, \mathbf{L} = \mathbf{l}\right)\\ &\quad +\mathbb{E}\left[Y_i(t) \mid W_i > w_j, W_j = w_j, \mathbf{X} = \mathrm{x}, \mathbf{L} = \mathbf{l}\right]\Pr\left(W_i > w_j \mid X_i = x_i, \mathbf{L} = \mathbf{l}\right)\Big)\\ &= \Pr\left(W_i < t \mid X_i = x_i, \mathbf{l}_i = \mathbf{l}_i\right)\Big)\\ &\quad +\mathbb{E}\left[Y_i(t) \mid W_i > w_j, W_j = w_j, \mathbf{X} = \mathrm{x}, \mathbf{L} = \mathbf{l}\right]\Pr\left(W_i > w_j \mid X_i = x_i, \mathbf{l}_i = \mathbf{l}_i\right)\Big)\\ &= \Pr\left(W_i \le t \mid X_i = x_i, \mathbf{l}_i = \mathbf{l}_i\right)\Big)\end{aligned}$$

since $\mathbb{E}\left[Y_i(t) \mid W_i > w_j, W_j = w_j, \mathbf{X} = \mathrm{x}, \mathbf{L} = \mathbf{l}\right] = 0$ when $t < w_j$

$$= F_i\left(w_j \mid x_i, l_i\right).$$

Proof of Corollary 1.

$$\begin{aligned}\mathbb{E}\left[Y_i\left(t; W_j\left(x_j\right), \mathrm{x}\right) \mid \mathbf{L}=\mathbf{l}\right] &= \mathbb{E}\left[\mathbb{E}\left[Y_i\left(t; W_j\left(x_j\right), \mathrm{x}\right) \mid \mathbf{L}=\mathbf{l}\right]\right] \\ &= \int_0^{\infty} \mathbb{E}\left[Y_i(t; u, \mathrm{x}) \mid W_j=u, \mathbf{X}=\mathrm{x}, \mathbf{L}=\mathbf{l}\right] dF_j\left(u \mid x_j, \mathbf{l}_i\right) \text{ by Assumption 1} \\ &= \int_0^{\infty} \mathbb{E}\left[Y_i(t) \mid W_j=u, \mathbf{X}=\mathrm{x}, \mathbf{L}=\mathbf{l}\right] dF_j\left(u \mid x_j, \mathbf{l}_i\right) \text{ by Lemma 3 and Assumption 4} \\ &= \mathbb{E}\left[Y_i(t) \mid \mathbf{X}=\mathrm{x}, \mathbf{L}=\mathbf{l}\right].\end{aligned}$$

Likewise, when $\mathbf{x} = (x_i, x_j)$ and $x'_j \neq x_j$,

$$\begin{aligned}\mathbb{E}\left[Y_i\left(t; W_j\left(x'_j\right), \mathrm{x} \mid \mathbf{L}=\mathbf{l}\right]\right] &= \mathbb{E}\left[\mathbb{E}\left[Y_i\left(t; W_j\left(x_j\right), \mathrm{x} \mid \mathbf{L}=\mathbf{l}\right]\right]\right] \\ &= \int_0^{\infty} \mathbb{E}\left[Y_i(t; u, \mathbf{X}) \mid W_j=u, \mathbf{X}=\mathrm{x}, \mathbf{L}=\mathbf{l}\right] dF_j\left(u \mid x'_j, \mathbf{l}_i\right) \text{ by Assumption 1} \\ &= \int_0^{\infty} \mathbb{E}\left[Y_i(t) \mid W_j=u, \mathbf{X}=\mathrm{x}, \mathbf{L}=\mathbf{l}\right] dF_j\left(u \mid x'_j, \mathbf{l}_i\right) \text{ by Lemma 3 and Assumption 4}\end{aligned}$$

□

Lemma 4.

When $\mathrm{SE}(t, w_j, x_j) = 0$, *then* $F_j(t|x_j) = F_j(t|1 - x_j)$ *and* $\mathbb{E}\left[Y_i(t) \mid X_i=1, X_j=x_j\right] = \mathbb{E}\left[Y_i(t) \mid X_i=0, X_j=x_j\right]$, *for all* $x_j \in \{0,1\}$ *and* $t \geq 0$.

When $\mathrm{SE}(t, w_j, x_j) = \mathrm{IE}(t, w_j, x_i) = 0$, *then* $\mathbb{E}\left[Y_i(t) \mid X_i=0, X_j=1\right] = \mathbb{E}\left[Y_i(t) \mid X_i=0, X_j=0\right]$.

When $\mathrm{SE}(t, w_j, x_j) = 0$ *and* $\mathrm{IE}(t, w_j, x_i) < 0$, *then* $\mathbb{E}\left[Y_i(t) \mid X_i=0, X_j=1\right] < \mathbb{E}\left[Y_i(t) \mid X_i=0, X_j=0\right]$.

Proof of Lemma 4.

First we prove $F_j(t|x_j) = F_j(t|1-x_j)$, for all $x_j \in \{0, 1\}$ when $\mathrm{SE}(t, w_j, x_j) = 0$.

$$\begin{aligned}F_j\left(t \mid x_j\right) &= \Pr\left(W_j\left(x_j\right) < t\right) = \Pr\left(T_j\left(w_i=\infty, x_i, x_j\right) < t\right) \text{ by the definition of } T_j\left(w_i, x_j, x_i\right) \\ &= \mathbb{E}\left[Y_j\left(t; w_i=\infty, x_j, x_i\right)\right] \text{ by the definition of } Y_j\left(u; w_i, x_j, x_i\right) \\ &= \mathbb{E}\left[Y_j\left(t; w_i=\infty, x'_j, x_i\right)\right] \text{ since } \mathrm{SE}\left(t, w_j, x_j\right) = 0 \\ &= \Pr\left(W_j\left(x'_j\right) < t\right) = F_j\left(t \mid x'_j\right).\end{aligned} \tag{11}$$

Second, we prove $\mathbb{E}\left[Y_i(t) \mid X_i=1, X_j=x_j\right] = \mathbb{E}\left[Y_i(t) \mid X_i=0, X_j=x_j\right]$ for all $x_j \in \{0,1\}$, if $\mathrm{SE}(t, w_j, x_j) = 0$.

$$\mathbb{E}\left[Y_i(t) \mid X_i = 1, X_j = x_j\right] = \int_0^\infty \mathbb{E}\left[Y_i(t) \mid W_j = u, X_i = 1, X_j = x_j\right] dF_j\left(u \mid x_j\right) \text{ by}$$

Assumption 1

$$\begin{aligned} &= \int_0^\infty \mathbb{E}\left[Y_i\left(t; u, x_i = 1, x_j\right)\right] dF_j\left(u \mid x_j\right) \text{ by Lemma 3} \\ &= \int_0^\infty \mathbb{E}\left[Y_i\left(t; u, x_i = 0, x_j\right)\right] dF_j\left(u \mid x_j\right) \text{ since SE}\left(t, w_j, x_j\right) = 0 \\ &= \mathbb{E}\left[Y_i(t) \mid X_i = 0, X_j = x_j\right]. \end{aligned} \tag{12}$$

Third, by (11), we prove $\mathbb{E}\left[Y_i(t) \mid X_i = 0, X_j = 1\right] = \mathbb{E}\left[Y_i(t) \mid X_i = 0, X_j = 0\right]$, if SE($t$, w_j, x_j) = IE(t, w_j, x_i) = 0.

$$\mathbb{E}\left[Y_i(t) \mid X_i = 0, X_j = 1\right] = \int_0^\infty \mathbb{E}\left[Y_i(t) \mid W_j = u, X_i = 0, X_j = 1\right] dF_j(u \mid 1) \text{ by}$$

Assumption 1

$$\begin{aligned} &= \int_0^\infty \mathbb{E}\left[Y_i\left(t; u, x_i = 0, x_j = 1\right)\right] dF_j(u \mid 1) \text{ by Lemma 3} \\ &= \int_0^\infty \mathbb{E}\left[Y_i\left(t; u, x_i = 0, x_j = 0\right)\right] dF_j(u \mid 1) \text{ since IE}\left(t, w_j, x_i\right) = 0 \\ &= \int_0^\infty \mathbb{E}\left[Y_i\left(t; u, x_i = 0, x_j = 0\right)\right] dF_j(u \mid 0) \text{by11} \\ &= \int_0^\infty \mathbb{E}\left[Y_i(t) \mid W_j = u, X_i = 0, x_j = 0\right)\right] dF_j(u \mid 0) \text{ by Lemma 3} \\ &= \mathbb{E}\left[Y_i(t) \mid X_i = 0, X_j = 0\right]. \end{aligned} \tag{13}$$

Fourth, by (11), we prove $\mathbb{E}\left[Y_i(t) \mid X_i = 0, X_j = 1\right] < \mathbb{E}\left[Y_i(t) \mid X_i = 0, X_j = 0\right]$, if SE($t$, w_j, x_j) = 0 and IE(t, w_j, x_i) < 0.

$$\mathbb{E}\left[Y_i(t) \mid X_i = 0, X_j = 1\right] = \int_0^\infty \mathbb{E}\left[Y_i(t) \mid W_j = u, X_i = 0, X_j = 1\right] dF_j(u \mid 1) \text{ by}$$

Assumption 1

$$\begin{aligned}
&= \int_0^\infty \mathbb{E}\left[Y_i\left(t; u, x_i = 0, x_j = 1\right)\right] dF_j(u \mid 1) \text{ by Lemma 3} \\
&< \int_0^\infty \mathbb{E}\left[Y_i\left(t; u, x_i = 0, x_j = 0\right)\right] dF_j(u \mid 1) \text{ since IE}\left(t, w_j, x_i\right) < 0 \qquad (14) \\
&= \int_0^\infty \mathbb{E}\left[Y_i\left(t; u, x_i = 0, x_j = 0\right)\right] dF_j(u \mid 0) \text{ by11} \\
&= \int_0^\infty \mathbb{E}\left[Y_i(t) \mid W_j = u, X_i = 0, x_j = 0\right)\Big] dF_j(u \mid 0) \text{ by Lemma 3} \\
&= \mathbb{E}\left[Y_i(t) \mid X_i = 0, X_j = 0\right].
\end{aligned}$$

□

Proof of Theorem 2.

Given the conclusions from (12) and (14), we have

$$\begin{aligned}
\text{DE}(t) &= \mathbb{E}\left[Y_i(t) \mid X_i = 1\right] - \mathbb{E}\left[Y_i(t) \mid X_i = 0\right] \\
&= \mathbb{E}\left[Y_i(t) \mid X_i = 1, X_j = 1\right] \Pr\left(X_j = 1 \mid X_i = 1\right) \\
&+ \mathbb{E}\left[Y_i(t) \mid X_i = 1, X_j = 0\right] \Pr\left(X_j = 0 \mid X_i = 1\right) \\
&- \mathbb{E}\left[Y_i(t) \mid X_i = 0, X_j = 1\right] \Pr\left(X_j = 1 \mid X_i = 0\right) - \mathbb{E}\left[Y_i(t) \mid X_i = 0, X_j = 0\right] \Pr\left(X_j = 0 \mid X_i = 0\right) \\
&= \mathbb{E}\left[Y_i(t) \mid X_i = 0, X_j = 1\right] \Pr\left(X_j = 1 \mid X_i = 1\right) \\
&+ \mathbb{E}\left[Y_i(t) \mid X_i = 0, X_j = 0\right] \Pr\left(X_j = 0 \mid X_i = 1\right) \\
&- \mathbb{E}\left[Y_i(t) \mid X_i = 0, X_j = 1\right] \Pr\left(X_j = 1 \mid X_i = 0\right) - \mathbb{E}\left[Y_i(t) \mid X_i = 0, X_j = 0\right] \Pr\left(X_j = 0 \mid X_i = 0\right)
\end{aligned}$$

by (12) in Lemma 4

$$\begin{aligned}
&= \mathbb{E}\left[Y_i(t) \mid X_i = 0, X_j = 1\right]\left[\Pr\left(X_j = 1 \mid X_i = 1\right) - \Pr\left(X_j = 1 \mid X_i = 0\right)\right] \\
&\quad + \mathbb{E}\left[Y_i(t) \mid X_i = 0, X_j = 0\right]\left[\Pr\left(X_j = 0 \mid X_i = 1\right)\right. \\
&\left. - \Pr\left(X_j = 0 \mid X_i = 0\right)\right] \\
&\quad = \mathbb{E}\left[Y_i(t) \mid X_i = 0, X_j = 1\right]\left[\Pr\left(X_j = 1 \mid X_i = 1\right)\right. \\
&\left. - \Pr\left(X_j = 1 \mid X_i = 0\right)\right] \qquad (15) \\
&\quad + \mathbb{E}\left[Y_i(t) \mid X_i = 0, X_j = 0\right]\left\{\left[1 - \Pr\left(X_j = 1 \mid X_i = 1\right)\right]\right. \\
&\left. - \left[1 - \Pr\left(X_j = 1 \mid X_i = 0\right)\right]\right\} \\
&\quad = \left\{\mathbb{E}\left[Y_i(t) \mid X_i = 0, X_j = 1\right] - \mathbb{E}\left[Y_i(t) \mid X_i = 0, X_j = 0\right]\right\} \\
&\cdot \left[\Pr\left(X_j = 1 \mid X_i = 1\right) - \Pr\left(X_j = 1 \mid X_i = 0\right)\right]
\end{aligned}$$

Note by (14) in Lemma 4, we have the first term at the last line of (15) being negative. The sign of DE(t) then depends only on the treatment assignment mechanism, which leads to the following conclusions for DE(t).

1. If the treatment assignment is positively correlated ($\Pr(X_i = c, X_j = c) > \Pr(X_i = c)\Pr(X_j = c)$ for $c \in \{0, 1\}$), we have:

$$\begin{aligned}
&\Pr\left(X_j = 1 \mid X_i = 1\right) - \Pr\left(X_j = 1 \mid X_i = 0\right) \\
&= \frac{\Pr\left(X_j = 1, X_i = 1\right)}{\Pr\left(X_i = 1\right)} - \frac{\Pr\left(X_j = 1, X_i = 0\right)}{\Pr\left(X_i = 0\right)} \\
&= \frac{\Pr\left(X_j = 1, X_i = 1\right)\Pr\left(X_i = 0\right) - \Pr\left(X_j = 1, X_i = 0\right)\Pr\left(X_i = 1\right)}{\Pr\left(X_i = 1\right)\Pr\left(X_i = 0\right)} \\
&= \frac{\Pr\left(X_j = 1, X_i = 1\right)\left[1 - \Pr\left(X_i = 1\right)\right] - \Pr\left(X_j = 1, X_i = 0\right)\Pr\left(X_i = 1\right)}{\Pr\left(X_i = 1\right)\Pr\left(X_i = 0\right)} \\
&= \frac{\Pr\left(X_j = 1, X_i = 1\right) - \Pr\left(X_j = 1, X_i = 1\right)\Pr\left(X_i = 1\right) - \Pr\left(X_j = 1, X_i = 0\right)\Pr\left(X_i = 1\right)}{\Pr\left(X_i = 1\right)\Pr\left(X_i = 0\right)} \\
&= \frac{\Pr\left(X_j = 1, X_i = 1\right) - \Pr\left(X_j = 1\right)\Pr\left(X_i = 1\right)}{\Pr\left(X_i = 1\right)\Pr\left(X_i = 0\right)} \\
&\geq 0
\end{aligned} \tag{16}$$

   Thus, DE(t) < 0.

2. If the treatment assignment is independent ($\Pr(X_i = c, X_j = c) = \Pr(X_i = c)\Pr(X_j = c)$ for $c \in \{0, 1\}$), then by similar arguments of (16), we have $\Pr(X_j = 1|X_i = 1) - \Pr(X_j = 1|X_i = 0) = 0$. Thus, DE($t$) = 0.

3. If the treatment assignment is negatively correlated ($\Pr(X_i = c, X_j = c) < \Pr(X_i = c)\Pr(X_j = c)$ for $c \in \{0, 1\}$), then by similar arguments of (16), we have $\Pr(X_j = 1|X_i = 1) - \Pr(X_j = 1|X_i = 0) < 0$. Thus, DE($t$) > 0.

When IE(t, w_j, x_i) = 0, following (13) and (15) in Lemma 4, we have $\mathbb{E}\left[Y_i(t) \mid X_i = 0, X_j = 1\right] = \mathbb{E}\left[Y_i(t) \mid X_i = 0, X_j = 0\right]$ and thus DE(t) = 0.

Similar arguments apply for $\mathrm{VE}_{\mathrm{AR}}(t)$. □

Proof of Theorem 3.

We evaluate the sign of VE_I^{net} by analyzing $\mathrm{SAR}_{00}(t) - \mathrm{SAR}_{10}(t)$.

$$\mathrm{VE}_I^{net}(t) = 1 - \frac{\mathrm{SAR}_{10}(t)}{\mathrm{SAR}_{00}(t)} = \frac{\mathrm{SAR}_{00}(t) - \mathrm{SAR}_{10}(t)}{\mathrm{SAR}_{00}(t)}$$

First, we analyze the sign of VE_I^{net} under a null true infectiousness effect, when the infection outcome is positively contagious and vaccine has a favorable effect prior to first infection through $h_0(u|1) = \varepsilon h_0(u|0)$, for $\varepsilon \in [0,1)$.

$$\begin{aligned}
&\mathrm{SAR}_{10}(t) - \mathrm{SAR}_{00}(t) \\
&= \mathbb{E}\left[Y_i(t) \mid T_j < t, T_i > T_j, X_i = 0, X_j = 1\right] - \mathbb{E}\left[Y_i(t) \mid T_j < t, T_i > T_j, X_i = 0, X_j = 0\right] \\
&= \frac{\int_0^t \mathbb{E}\left[Y_i(t) \mid W_j = u, W_i > u, \mathbf{X} = (0,1)\right]\left(1 - F_i(u \mid 0)\right) dF_j(u \mid 1)}{\Pr\left(W_j < t, W_i > W_j \mid \mathbf{X} = (0,1)\right)} \\
&- \frac{\int_0^t \mathbb{E}\left[Y_i(t) \mid W_j = u, W_i > u, \mathbf{X} = (0,0)\right]\left(1 - F_i(u \mid 0)\right) dF_j(u \mid 0)}{\Pr\left(W_j < t, W_i > W_j \mid \mathbf{X} = (0,0)\right)}
\end{aligned}$$

by applying the law of total probability

$$\begin{aligned}
&= \int_0^t \mathbb{E}\left[Y_i(t) \mid W_j = u, W_i > u, \mathbf{X} = (0,1)\right] \frac{\left(1 - F_i(u \mid 0)\right) dF_j(u \mid 1)}{\int_0^t \left(1 - F_i(v \mid 0)\right) dF_j(v \mid 1)} \\
&- \int_0^t \mathbb{E}\left[Y_i(t) \mid W_j = u, W_i > u, \mathbf{X} = (0,0)\right] \frac{\left(1 - F_i(u \mid 0)\right) dF_j(u \mid 0)}{\int_0^t \left(1 - F_i(v \mid 0)\right) dF_j(v \mid 0)} \\
&= \int_0^t \mathbb{E}\left[Y_i(t) \mid W_j = u, W_i > u, \mathbf{X} = (0,0)\right] \\
&\left[\frac{\left(1 - F_i(u \mid 0)\right) dF_j(u \mid 1)}{\int_0^t \left(1 - F_i(v \mid 0)\right) dF_j(v \mid 1)} - \frac{\left(1 - F_i(u \mid 0)\right) dF_j(u \mid 0)}{\int_0^t \left(1 - F_i(v \mid 0)\right) dF_j(v \mid 0)}\right].
\end{aligned} \tag{17}$$

By IE(t, w_j, 0) = 0 and Lemma 3 To ease the notation in Equation (17), we denote $\mathbb{E}\left[Y_i(t) \mid W_j = u, W_i > u, \mathbf{X} = (0,0)\right] = k(u)$. Denote $g(u \mid 1) = \frac{\left(1 - F_i(u \mid 0)\right) dF_j(u \mid 1)}{\int_0^t \left(1 - F_i(v \mid 0)\right) dF_j(v \mid 1)}$ and $g(u \mid 0) = \frac{\left(1 - F_i(u \mid 0)\right) dF_j(u \mid 0)}{\int_0^t \left(1 - F_i(v \mid 0)\right) dF_j(v \mid 0)}$, and $G(u \mid 1) = \int_0^u g(s \mid 1) ds$ and $G(u \mid 0) = \int_0^u g(s \mid 0) ds$. Then by integration by parts, (17) can be re-written as follows:

$$\begin{aligned}
\mathrm{SAR}_{10}(t) - \mathrm{SAR}_{00}(t) &= \int_0^t k(u)[g(u \mid 1) - g(u \mid 0)] du \\
&= k(u)[G(u \mid 1) - G(u \mid 0)]_0^t - \int_0^t (G(u \mid 1) - G(u \mid 0)) dk(u).
\end{aligned}$$

By their definitions, we have $G(0|1) - G(0|0) = 0$ and $G(t|1) - G(t|0) = 0$, and thus $k(u)[G(u \mid 1) - G(u \mid 0)]_0^t = 0$. In other words, the sign of $\mathrm{SAR}_{10}(t) - \mathrm{SAR}_{00}(t)$ only depends on the sign of $G(u|1) - G(u|0)$ and $dk(u)$ for all $u > 0$. First, we can show that $dk(u) < 0$ for $0 \le u < t$. For $0 \le u < u' < t$, we have

$$
\begin{aligned}
k(u) &= \frac{\mathbb{E}\left[Y_i(t) \mid W_j = u, \mathbf{X} = (0,0)\right] - F_i(u \mid 0)}{1 - F_i(u \mid 0)} \text{ by Theorem 1} \\
&> \frac{\mathbb{E}\left[Y_i(t) \mid W_j = u', \mathbf{X} = (0,0)\right] - F_i(u \mid 0)}{1 - F_i(u \mid 0)} \text{ by } \mathrm{CE}(t, u, u', (0,0)) > 0 \\
&= \frac{\mathbb{E}\left[Y_i(t) \mid W_j = u', \mathbf{X} = (0,0)\right] - F_i(u' \mid 0) + F_i(u' \mid 0) - F_i(u \mid 0)}{1 - F_i(u \mid 0)} \\
&= \frac{k(u')\left(1 - F_i(u' \mid 0)\right) + F_i(u' \mid 0) - F_i(u \mid 0)}{1 - F_i(u \mid 0)} \text{ by Theorem 1} \\
&\geq \frac{k(u')\left(1 - F_i(u' \mid 0)\right) + \left(F_i(u' \mid 0) - F_i(u \mid 0)\right)k(u')}{1 - F_i(u \mid 0)} \text{ by } k(u') \leq 1 \\
&= \frac{k(u')\left(1 - F_i(u \mid 0)\right)}{1 - F_i(u \mid 0)} = k(u').
\end{aligned}
\tag{18}
$$

Next, we analyze the property of $G(u|1) - G(u|0)$ for $\forall u > 0$. Denote $H_0(u) = \int_0^u h_0(s \mid 0)ds$. Given $h_0(u|1) = \varepsilon h_0(u|0)$ with $\varepsilon \in [0,1)$, we can write out $G(u|0)$ and $G(u|1)$ in terms of $h_0(u|0)$ as follows.

$$
\begin{aligned}
G(s \mid 1) &= \frac{\int_0^s \left(1 - F_i(u \mid 0)\right) dF_j(u \mid 1)}{\int_0^t \left(1 - F_i(v \mid 0)\right) dF_j(v \mid 1)} = \frac{\int_0^s \varepsilon \cdot h_0(u \mid 0) e^{-\varepsilon \cdot H_0(u)} e^{-H_0(u)} du}{\int_0^t \varepsilon \cdot h_0(v \mid 0) e^{-\varepsilon \cdot H_0(v)} e^{-H_0(v)} dv} \\
&= \frac{\int_0^s \varepsilon \cdot h_0(u \mid 0) e^{-(\varepsilon+1) \cdot H_0(u)} du}{\int_0^t \varepsilon \cdot h_0(v \mid 0) e^{-(\varepsilon+1) \cdot H_0(v) dv}} = \frac{1 - e^{-(\varepsilon+1) H_0(s)}}{1 - e^{-(\varepsilon+1) H_0(t)}} \\
G(s \mid 0) &= \frac{\int_0^s \left(1 - F_i(u \mid 0)\right) dF_j(u \mid 1)}{\int_0^t \left(1 - F_i(v \mid 0)\right) dF_j(v \mid 1)} = \frac{1 - e^{-2H_0(s)}}{1 - e^{-2H_0(t)}}
\end{aligned}
\tag{19}
$$

From (19), we observe that $G(s|1)$ and $G(s|0)$ only differ by the terms in front of H_0. Treat $G(s|1)$ and $G(s|0)$ as functions of ε, and we can re-express them as $G(\varepsilon) = \frac{1 - e^{-(\varepsilon+1)H_0(s)}}{1 - e^{-(\varepsilon+1)H_0(t)}}$ and $G(1) = \frac{1 - e^{-2H_0(s)}}{1 - e^{-2H_0(t)}}$, given $\varepsilon < 1$. Then, if $G(\varepsilon)$ is a decreasing function of ε, we have $G(u|1) - G(u|0) \leq 0$.

$$
\begin{aligned}
&\frac{\partial}{\partial \varepsilon} G(\varepsilon) \\
&= \frac{H_0(u) e^{-(\varepsilon+1)H_0(u)} \left[1 - e^{-(\varepsilon+1)H_0(t)}\right] - H_0(t) e^{-(\varepsilon+1)H_0(t)} \left[1 - e^{-(\varepsilon+1)H_0(u)}\right]}{\left[1 - e^{-(\varepsilon+1)H_0(t)}\right]^2}
\end{aligned}
\tag{20}
$$

Divide the numerator of (20) by a positive constant $H_0(t)H_0(u)e^{-(\varepsilon+1)\left[H_0(u)+H_0(t)\right]}$. We then have if $\frac{e^{(\varepsilon+1)H_0(u)}-1}{H_0(u)} \le \frac{e^{(\varepsilon+1)H_0(t)}-1}{H_0(t)}$ for $u < t$, then $G(u|1) - G(u|0) \le 0$. Treat $\frac{e^{(\varepsilon+1)H_0(t)}-1}{H_0(t)}$ as a function of u, given $0 \le u < t$. We have,

$$\begin{aligned}\frac{\partial}{\partial u}\frac{e^{(\varepsilon+1)H_0(u)}-1}{H_0(u)} &= \frac{(\varepsilon+1)H_0(u)e^{(\varepsilon+1)H_0(u)} - e^{(\varepsilon+1)H_0(u)} + 1}{\left[H_0(u)\right]^2} \\ &= \frac{(\varepsilon+1)H_0(u) - 1 + e^{-(\varepsilon+1)H_0(u)}}{\left[H_0(u)\right]^2 e^{(\varepsilon+1)H_0(u)}} \text{ by } e^{-(\varepsilon+1)H_0(u)} \ge 1 - (\varepsilon \\ &+ 1)H_0(u) \\ &\ge 0.\end{aligned} \tag{21}$$

Combining (20) and (21), we have $G(u|1) - G(u|0) \le 0$.

In summary, we can see that

$$\mathrm{SAR}_{10}(t) - \mathrm{SAR}_{00}(t) = k(u)[G(u \mid 1) - G(u \mid 0)]\Big|_0^t - \int_0^t (G(u \mid 1) - G(u \mid 0))dk(u) < 0$$

Thus, $\mathrm{VE}_I^{net}(t) = 1 - \frac{\mathrm{SAR}_{10}(t)}{\mathrm{SAR}_{00}(t)} = \frac{\mathrm{SAR}_{00}(t) - \mathrm{SAR}_{10}(t)}{\mathrm{SAR}_{00}(t)} > 0.$

Next, we analyze the sign of $\mathrm{VE}_I^{net}(t)$ under a null true susceptibility effect.

$$\begin{aligned}&\mathrm{SAR}_{10}(t) - \mathrm{SAR}_{00}(t) \\ &= \int_0^t \mathbb{E}\Big[Y_i(t) \mid W_j = u, W_i > u, \mathbf{X} = (0,1)\Big]\frac{\left(1 - F_i(u \mid 0)\right)dF_j(u \mid 1)}{\int_0^t \left(1 - F_i(v \mid 0)\right)dF_j(v \mid 1)} \\ &- \int_0^t \mathbb{E}\Big[Y_i(t) \mid W_j = u, W_i > u, \mathbf{X} = (0,0)\Big] - \frac{\left(1 - F_i(u \mid 0)\right)dF_j(u \mid 0)}{\int_0^t \left(1 - F_i(v \mid 0)\right)dF_j(v \mid 0)}\Bigg]\end{aligned}$$

by (17)

$$= \int_0^t \left\{ \mathbb{E}\left[Y_i(t) \mid W_j = u, W_i > u, \mathbf{X} = (0,1) \right] - \mathbb{E}\left[Y_i(t) \mid W_j = u, W_i > u, \mathbf{X} = (0,0) \right] \right\} \cdot \frac{\left(1 - F_i(u \mid 0)\right) dF_j(u \mid 0)}{\int_0^t \left(1 - F_i(v \mid 0)\right) dF_j(v \mid 0)}$$

by $\mathrm{SE}(t, w_j, x_j) = 0$ and (11)

$$= \int_0^t \left\{ \frac{\mathbb{E}\left[Y_i(t) \mid W_j = u, \mathbf{X} = (0,1) \right]}{\Pr\left(W_i > u \mid W_j = u, \mathbf{X} = (0,1) \right)} - \frac{\mathbb{E}\left[Y_i(t) \mid W_j = u, \mathbf{X} = (0,0) \right]}{\Pr\left(W_i > u \mid W_j = u, \mathbf{X} = (0,0) \right)} \right\} \frac{\left(1 - F_i(u \mid 0)\right) dF_j(u \mid 0)}{\int_0^t \left(1 - F_i(v \mid 0)\right) dF_j(v \mid 0)}$$

$$= \int_0^t \left\{ \frac{\mathbb{E}\left[Y_i(t) \mid W_j = u, \mathbf{X} = (0,1) \right]}{\Pr\left(W_i > u \mid X_i = 0 \right)} - \frac{\mathbb{E}\left[Y_i(t) \mid W_j = u, \mathbf{X} = (0,0) \right]}{\Pr\left(W_i > u \mid X_i = 0 \right)} \right\} \frac{\left(1 - F_i(u \mid 0)\right) dF_j(u \mid 0)}{\int_0^t \left(1 - F_i(v \mid 0)\right) dF_j(v \mid 0)}$$

by Assumption 1

$$= \int_0^t \left\{ \mathbb{E}\left[Y_i(t) \mid W_j = u, \mathbf{X} = (0,1) \right] - \mathbb{E}\left[Y_i(t) \mid W_j = u, \mathbf{X} = (0,0) \right] \right\} \frac{\left(1 - F_i(u \mid 0)\right)}{\Pr\left(W_i > u \mid X_i = 0 \right)} dF_j(u \mid 0) \cdot \frac{1}{\int_0^t \left(1 - F_i(v \mid 0)\right) dF_j(v \mid 0)}$$

$$= \int_0^t \frac{\mathrm{IE}(t, u, 0)\left(1 - F_i(u \mid 0)\right)}{\Pr\left(W_i > u \mid X_i = 0 \right)} dF_j(u \mid 0) \cdot \frac{1}{\int_0^t \left(1 - F_i(v \mid 0)\right) dF_j(v \mid 0)}$$

Thus, $\mathrm{VE}_I^{net}(t)$ has the same sign as the true infectiousness effect, when the true susceptibility effect is null.

Third, we analyze the sign of $\mathrm{VE}_I^{net}(t)$ in the case of no contagion, when the true susceptibility effect is beneficial. First, $\mathrm{CE}\left(t, w_j, w_j', \mathbf{0}\right) = 0$ for all $0 < w_j < w_j'$ implies $\mathrm{IE}(t, w_j, 0) = 0$.

$$\mathrm{IE}\left(t, w_j, x_i\right) = \mathbb{E}\left[Y_i\left(t; w_j, x_i, x_j = 1\right) - Y_i\left(t; w_j, x_i, x_j = 0\right) \right] = \mathbb{E}\left[1\left\{ W_i\left(x_i\right) < t \right\} - 1\left\{ W_i\left(x_i\right) < t \right\} \right] = 0$$

Following the same proof for the first case except replacing the second line of (18) by an equal sign, we know $\mathrm{VE}_I^{net}(t) > 0$. □

Proof of Theorem 4.

Given $h_0^i(t \mid 0) = 0$, we have $F_i(s \mid 0) = 1 - e^{-\int_0^s h_0^i(u \mid 0) du} = 0$ for $W_i(0)$.

$$\mathbb{E}\left[Y_i\left(t; Y_j\left(x'_j\right), \left(0, x_j\right)\right) \mid h_0^i(t \mid 0) = 0\right] = \mathbb{E}\left[Y_i\left(t; Y_j\left(x'_j\right), \left(0, x_j\right)\right) \mid W_i(0) = \infty\right]$$

by $F_i(s|x_i) = 0$ for $\forall s > 0$

$$= \mathbb{E}\left[Y_i\left(t; 1\left\{W_j\left(x'_j\right) < t\right\}, \left(0, x_j\right)\right) \mid W_i(0) = \infty\right]$$

given $Y_i\left(x'_j\right) = \left\{T_j\left(x'_j\right) < t\right\}$ and $T_j\left(x'_j\right) = W_j\left(x'_j\right)$ when $W_i(0) = \infty$

$$\begin{aligned} &= \mathbb{E}\left[Y_i\left(t; W_j\left(x'_j\right), \left(0, x_j\right)\right) \mid W_i(0) = \infty\right] \\ &= \mathbb{E}\left[Y_i\left(t; W_j\left(x'_j\right), \left(0, x_j\right)\right) \mid h_0^i(s \mid 0) = 0\right] \end{aligned} \tag{22}$$

Thus, by the definition of $\mathrm{VE}_I(t)$ and $\mathrm{IE}(t, x_i)$, we have:

$$\begin{aligned} \mathrm{VE}_I(t) &= \mathbb{E}\left[Y_i\left(t; Y_j(1), (0, 1)\right) - Y_i\left(t; Y_j(1), (0, 0)\right) \mid h_0^i(s \mid 0) = 0\right] \\ &= \mathbb{E}\left[Y_i\left(t; W_j(1), (0, 1)\right) - Y_i\left(t; W_j(1), (0, 0)\right) \mid h_0^i(s \mid 0) = 0\right] = \mathrm{IE}\left(t, 0 \mid h_0^i(s \mid 0) = 0\right) \end{aligned}$$

Thus, VE_I is equivalent to the natural infectiousness effect under the asymmetric partnership. □

Proof of Theorem 5.

Given $h_0^i(t \mid 0) = 0$, we have $\mathbb{E}[Yi(t; wj, \mathrm{x}) \mid \mathbf{L} = 1]$.

$$\begin{aligned} \mathrm{VE}_C(t) &= \mathbb{E}\left[Y_i\left(t; Y_j(1), (0, 0)\right)\right] - \mathbb{E}\left[Y_i\left(t; Y_j(0), (0, 0)\right)\right] \\ &= \mathbb{E}\left[Y_i\left(t; W_j(1), (0, 0)\right)\right] - \mathbb{E}\left[Y_i\left(t; W_j(0), (0, 0)\right)\right] \end{aligned}$$

by Equation (22)

$$= \int_0^\infty \mathbb{E}\left[Y_i\left(t; w_j, (0, 0)\right)\right] dF_j\left(w_j \mid 1\right) - \int_0^\infty \mathbb{E}\left[Y_i\left(t; w_j, (0, 0)\right] dF_j\left(w_j \mid 0\right)\right.$$

by Corollary 1

$$= \int_0^\infty \left\{F_i\left(w_j \mid 0\right) + \left(1 - F_i\left(w_j \mid 0\right)\right) \mathbb{E}\left[Y_i(t) \mid W_i > w_j, W_j = w_j, \mathbf{X} = (0, 0)\right]\right\} d\left(F_j\left(w_j \mid 1\right) - F_j\left(w_j \mid 0\right)\right)$$

by Theorem 1

$$= \int_0^\infty \mathbb{E}\left[Y_i(t) \mid W_i > w_j, W_j = w_j, \mathbf{X} = (0, 0)\right] d\left(F_j\left(w_j \mid 1\right) - F_j\left(w_j \mid 0\right)\right)$$

by $F_i(t|x_i) = 0$ for $\forall t > 0$

$$= \int_0^t \mathbb{E}\left[Y_i(t) \mid W_i > w_j, W_j = w_j, \mathbf{X} = (0,0)\right] d\left(F_j\left(w_j \mid 1\right) - F_j\left(w_j \mid 0\right)\right)$$

since $\mathbb{E}\left[Y_i(t) \mid W_i > w_j, W_j = w_j, \mathbf{X} = (0,0)\right] = 0$ for $w_j > t$

$$= \int_0^t k\left(w_j\right) d\left(F_j\left(w_j \mid 1\right) - F_j\left(w_j \mid 0\right)\right)$$

by the definition of k(u) in the proof of Theorem 3

$$= k\left(w_j\right)\left[F_j\left(w_j \mid 1\right) - F_j\left(w_j \mid 0\right)\right]\Big|_0^t - \int_0^t F_j\left(w_j \mid 1\right) - F_j\left(w_j \mid 0\right) dk\left(w_j\right)$$

by integration by parts

By the definition of $k(u)$ and $F_j(u|x_j)$, we know $k(t) = 0$ and $F_j(0|1) - F_j(0|0) = 0$, and thus $k\left(w_j\right)\left[F_j\left(w_j \mid 1\right) - F_j\left(w_j \mid 0\right)\right]\Big|_0^t = 0$ If $SE(t, w_j, x) > 0$, we have $F_j(w_j|1) - F_j(w_j|0) > 0$. If $CE(t, u, u', (0, 0)) > 0$ for $0 \le u < u' < t$, $dk(u) < 0$ as shown in the the proof of Theorem 3. Thus, we have the following conclusions.

When $SE(t, w_j, x) > 0$, $VE_C(t)$ has the opposite sign as $CE(t, u', u, (0, 0))$.

If $SE(t, w_j, x) = 0$ and $CE(t, u, u', (0, 0)) > 0$, we have $VE_C(t) = 0$. □

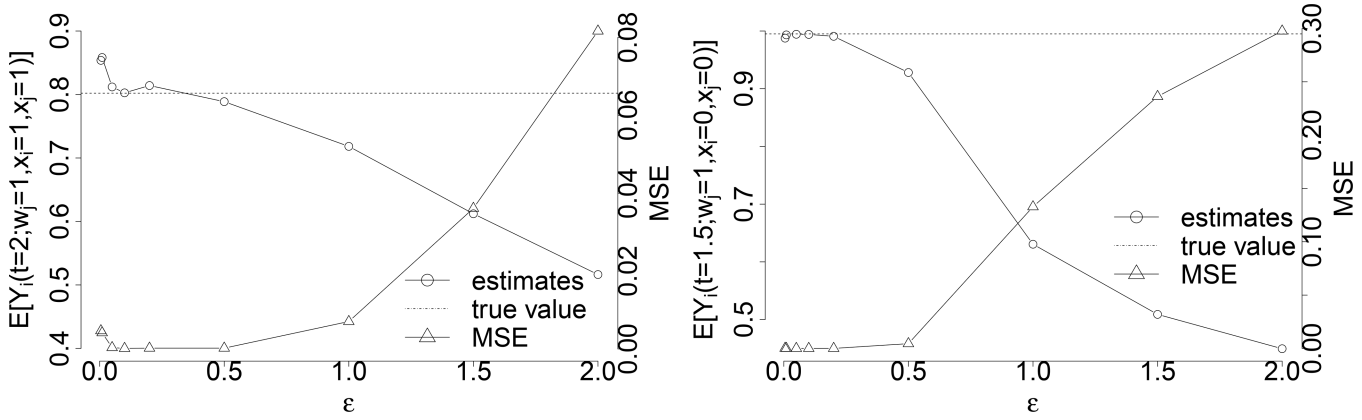

Figure 5:

The choice of ϵ in the estimation of $\mathbb{E}\left[Y_i\left(t; x_i, x_j, w_j\right)\right]$ with sample size $n = 100,000$ under the constant hazards $\alpha(t) = 0.2$, $\gamma(t) = 10$ and coefficients $e^{\beta_0} = e^{\beta_1} = 0.2$ and $e^{\sigma} = 0.5$. Figure on the left shows the estimation of $\mathbb{E}\left[Y_i\left(t = 2; x_i = 1, x_j = 1, s = 1\right)\right]$ and its corresponding MSE under different choices of ϵ, and Figure on the right shows the estimation of $\mathbb{E}\left[Y_i\left(t = 1.5; x_i = 0, x_j = 0, s = 1\right)\right]$ and its corresponding MSE under difference choices of ϵ.

B: Statistical estimation

B.1 Statistical estimation for the controlled potential outcomes in Theorem 1

In Theorem 1, for $t < w_j$, the estimation of $\mathbb{E}[Yi(t; wj, \mathrm{x}) \mid \mathbf{L} = 1]$ is achieved by the estimation of $F_I(w_j|x_i, \mathbf{l}_i)$ by Lemma 1, which follows the standard technique of estimating distribution of time-to-event data in competing risks. For $t \geq w_j$, the estimation of $\mathbb{E}\left[Y_i\left(t; w_j, \mathrm{x}\right) \mid \mathbf{L} = 1\right]$ is achieved by the estimation of $F_I(w_j|x_i, \mathbf{l}_i)$ by Lemma 1 and the estimation of $\mathbb{E}\left[Y_i(t) \mid T_i \geq w_j, T_j = w_j, \mathbf{X} = \mathbf{X}, \mathbf{L} = 1\right]$. Let ϵ be a small positive number, then

$$\mathbb{E}\left[Y_i(t) \middle| Ti \geq wj, Tj = wj, \mathrm{x} = x, L = 1\right] = \lim \varepsilon \to 0 \mathbb{E}\left[Yi(t) \middle| wj - \varepsilon < Tj < wj + \mathbf{X} = x, \mathbf{L} = 1\right].$$

Therefore, we estimate $\mathbb{E}\left[Y_i(t) \mid T_i \geq w_j, T_j = w_j, \mathbf{X} = \mathrm{x}, \mathbf{L} = \mathbf{l}\right]$ by averaging $Y_i(t)$ among observations when T_j falls into a narrow region around w_j under $\mathbf{X} = \mathbf{x}$ and $\mathbf{L} = \mathbf{l}$. With finite samples of observations, if ϵ is chosen too small, sample size for the estimation becomes smaller and variance gets bigger; if the ϵ is chosen too big, the selected observations no longer approximate $T_j = w_j$ well enough so that the estimation is more biased. The ϵ should be chosen to minimize the MSE of the estimation.

We choose $\epsilon = 0.1$ in the estimations of controlled potential outcomes in Figure 3 when $t \geq w_j$ with sample size $N = 100,000$ (under the constant hazard scenario $\alpha(t) = 0.2$ and $\gamma(t) = 10$ with beneficial susceptibility effect $\beta_1 = 0.3$ and infectiousness effect $\beta_2 = 0.5$), as it gives the smallest (or almost smallest) MSE for most observational times under different treatments and partner's infection time. Figure 5 illustrates the estimations of $\mathbb{E}\left[Y_i\left(2; x_i = 1, x_j = 1, w_j = 1\right)\right]$ and $\mathbb{E}\left[Y_i\left(1.5; x_i = 0, x_j = 0, w_j = 1\right)\right]$ as well as their MSEs for the choice of ϵ among $\epsilon \in \{0.005, 0.01, 0.05, 0.1, 0.2, 0.5, 1, 1.5, 2\}$, and $\epsilon = 0.1$ gives the smallest MSE for the estimation.

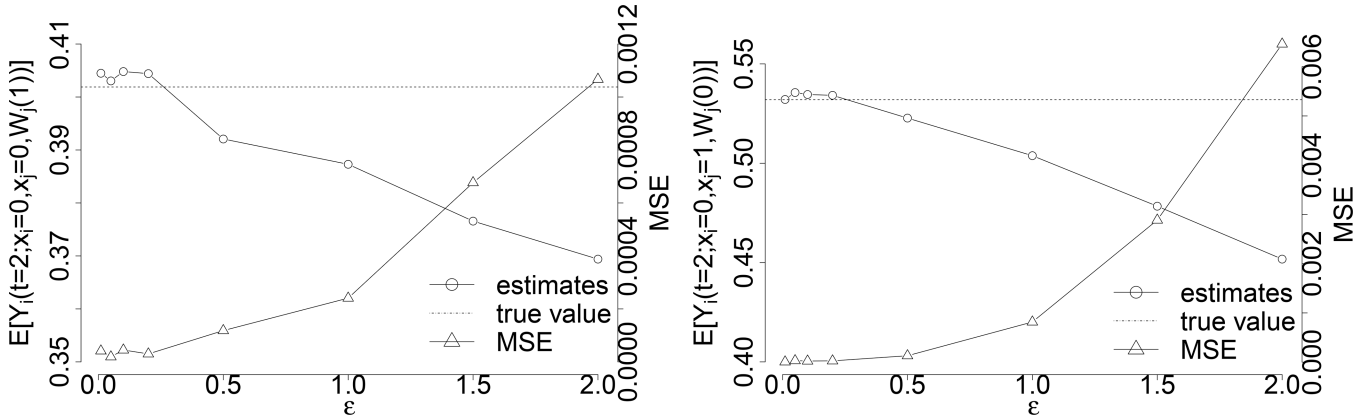

Figure 6:

The choice of ε in the estimation of $\mathbb{E}\left[Y_i\left(t; x_i, x_j, W_j\left(x'_j\right)\right)\right]$ with sample size $n = 100,000$ under the constant hazards $\alpha(t) = 0.2$, $\gamma(t) = 10$ and coefficients $e^{\beta_0} = e^{\beta_1} = 0.2$ and $e^{\sigma} = 0.5$. Figure on the left shows the estimation of $\mathbb{E}\left[Y_i\left(t = 2; 0, 0, W_j(1)\right)\right]$ and its corresponding MSE under different choices of ε, and Figure on the right shows the estimation of $\mathbb{E}\left[Y_i\left(t = 2; 0, 1, W_j(0)\right)\right]$ and its corresponding MSE under difference choices of ϵ.

B.2 Statistical estimation for the natural potential outcomes in Corollary 1

From Corollary 1, $\mathbb{E}\left[Y_i\left(t; x_i, x_j, W_j\left(x_j\right)\right)\right]$ can be estimated by the average of $Y_i(t)$ when $\mathbf{X} = \mathbf{x}$.

For the identification of cross-world natural potential outcomes when $x'_j \neq x_j$, $\mathbb{E}\left[Y_i\left(t; x_i, x_j, W_j\left(x'_j\right)\right)\right]$ is estimated with the help of the estimation of $F_j\left(w_j \mid x'_j, \mathbf{l}_j\right)$ by Lemma 1 and the estimation of $\mathbb{E}\left[Y_i\left(t; w_j, \mathrm{x}\right) \mid \mathbf{L} = \mathbf{l}\right]$ in Theorem 1, which requires a proper choosing of ε again.

We illustrate examples of estimating cross-world natural potential outcomes of $\mathbb{E}\left[Y_i\left(t = 2; 0, 0, W_j(1)\right)\right]$ and $\mathbb{E}\left[Y_i\left(t = 2; 0, 1, W_j(0)\right)\right]$ with sample size $N = 1,000,000$ under the constant hazard scenario ($\alpha(t) = 0.2$ and $\gamma(t) = 10$) with beneficial susceptibility effect ($\beta_1 = 0.3$) and infectiousness effect ($\beta_2 = 0.5$). We show their estimations as well as the MSEs under the choice among $\epsilon \in \{0.005, 0.01, 0.05, 0.1, 0.2, 0.5, 1, 1.5, 2\}$ in Figure 6, and $\epsilon = 0.1$ gives the smallest MSE for the estimations.

B.3 Covariate adjustment for controlled and natural potential infection outcomes in Equations (4)–(5)

For the adjustment of covariates in Equations (4)–(5), the estimation is achieved by estimating (controlled or natural) potential outcomes by Theorem 1 and Corollary 1, and then integrate it over the estimated empirical distribution of the covariates.

We approximate the joint distribution of covariates $G(\mathbf{l})$ empirically by dividing the space of $\mathbf{L}$ into small bins of size $\Delta \times \Delta$. The probability of $\mathbf{L}$ in one bin centered around (c_i, c_j) is estimated by $\Pr\left(c_i - \frac{\Delta}{2} < L_i < c_i + \frac{\Delta}{2}, c_j - \frac{\Delta}{2} < L_j < c_j + \frac{\Delta}{2}\right) = \frac{1}{N}\sum_i 1$. The size of Δ should be $\left\{c_i - \frac{\Delta}{2} < L_i < c_i + \frac{\Delta}{2}, c_j - \frac{\Delta}{2} < L_j < c_j + \frac{\Delta}{2}\right\}$ chosen to minimize the MSE of the estimations. Within each bin centered, for example the one around (c_i, c_j), we estimate $\mathbb{E}\left[Y_i\left(t; w_j, \mathrm{x}\right) \mid L_i = c_i, L_j = c_j\right]$ and $\mathbb{E}\left[Y_i\left(t; \mathbf{X}, W_j\left(x_j\right)\right) \mid L_i = c_i, L_j = c_j\right]$ by Theorem 1 and in Corollary 1, respectively. Finally, we integrate $\mathbb{E}\left[Y_i\left(t; w_j, \mathrm{x}\right) \mid L_i = c_i, L_j = c_j\right]$ and $\mathbb{E}\left[Y_i\left(t; W_j\left(x_j\right), \mathrm{x}\right) \mid L_i = c_i, L_j = c_j\right]$ over the estimated empirical distribution of $G(\mathbf{l})$ by:

$$\mathbb{E}\left[Y_i\left(t; w_j, \mathrm{x}\right)\right] = \sum_{c_i, c_j} \mathbb{E}\left[Y_i\left(t; w_j, \mathrm{x}\right) \mid L_i = c_i, L_j = c_j\right] \Pr\left(c_i - \frac{\Delta}{2} < L_i < c_i + \frac{\Delta}{2}, c_j - \frac{\Delta}{2} < L_j < c_j + \frac{\Delta}{2}\right)$$

$$\mathbb{E}\left[Y_i\left(t; W_j\left(x_j\right), \mathrm{x}\right)\right] =$$

$$\sum_{c_i, c_j} \mathbb{E}\left[Y_i\left(t; W_j\left(x_j\right), \mathrm{x}\right) \mid L_i = c_i, L_j = c_j\right] \Pr\left(c_i - \frac{\Delta}{2} < L_i < c_i + \frac{\Delta}{2}, c_j - \frac{\Delta}{2} < L_j < c_j + \frac{\Delta}{2}\right)$$

We illustrate the estimation of $\mathbb{E}\left[Y_i\left(t = 2; x_i = 1, x_j = 1, w_j = 1\right)\right]$ and $\mathbb{E}\left[Y_i\left(t = 2; x_i = 0, x_j = 0, W_j(0)\right)\right]$ with one covariate for each individual, so $\mathbf{L} = (L_i, L_j)$, with sample size $n = 1,000,000$ under the constant hazards $\alpha(t) = 0.2$, $\gamma(t) = 10$ and coefficients $e^{\beta_0} = e^{\beta_1} = 0.2$ and $e^{\sigma} = 0.5$. In our simulation, the covariates are generated by

$$\begin{pmatrix} L_i \\ L_j \end{pmatrix} \sim \mathrm{Normal}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, v\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$$

so that the majority of them fall into (−4,4). Therefore, we separate the covariates space into bins from −4 to 4 by Δ as well as the 4 left regions at the corners. Specifically, the space of (L_i, L_j) are separated into bins of $\left(c_i - \frac{\Delta}{2}, c_i + \frac{\Delta}{2}\right] \times \left(c_j - \frac{\Delta}{2}, c_j + \frac{\Delta}{2}\right]$, where $c_j, c_j \in \left\{-4 + \frac{\Delta}{2}, -4 + \frac{3\Delta}{2}, \ldots, 4 - \frac{3\Delta}{2}, 4 - \frac{\Delta}{2}\right\}$, as well as $(-\infty, -4] \times (-\infty, -4]$, $(-\infty, -4] \times (4, \infty)$, $(4, \infty) \times (-\infty, -4]$, and $(4, \infty) \times (4, \infty)$ at the corners.

We show the estimations of $\mathbb{E}\left[Y_i\left(t = 2; x_i = 1, x_j = 1, w_j = 1\right)\right]$ and $\mathbb{E}\left[Y_i\left(t = 2; x_i = 0, x_j = 0, W_j(0)\right)\right]$ as well as MSE under the choice among $\Delta \in \{0.005, 0.01, 0.05, 0.1, 0.2, 0.5, 1, 1.5, 2\}$ in Figure 7, and $\Delta = 0.1$ gives the smallest MSE for the estimations.

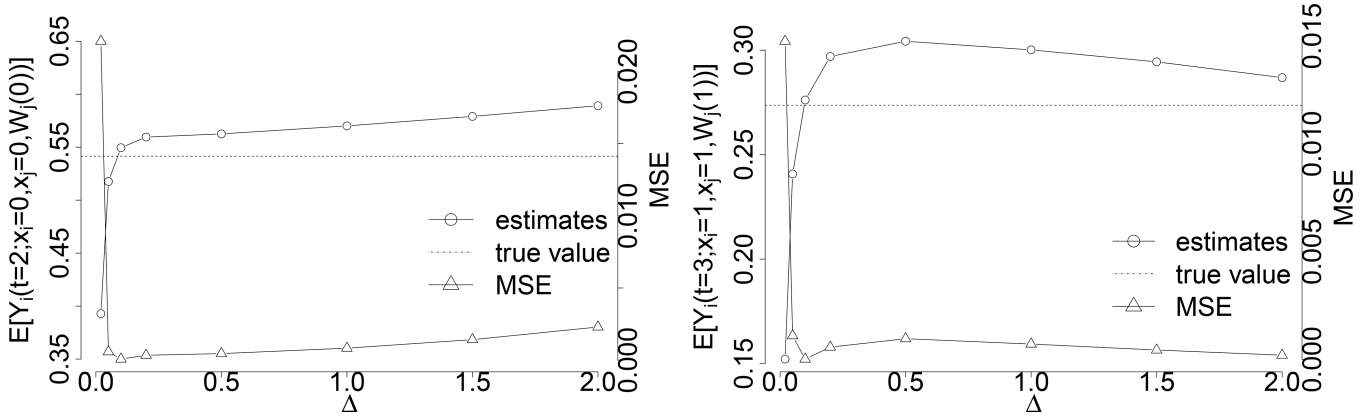

Figure 7:

The choice of Δ in the estimation of natural potential outcomes with sample size $n = 1,000,000$ under the constant hazards $\alpha(t) = 0.2$, $\gamma(t) = 10$ and coefficients $e^{\beta_0} = e^{\beta_1} = 0.2$ and $e^{\sigma} = 0.5$. Figure on the left shows the estimation of $\mathbb{E}\left[Y_i\left(2; 0, 0, W_j(0)\right)\right]$ and its corresponding MSE under different choices of Δ among $\Delta \in \{0.02, 0.05, 0.1, 0.2, 0.5, 1, 1.5, 2\}$, and Figure on the right shows the estimation of $\mathbb{E}\left[Y_i\left(3; x_i = 1, x_j = 1, W_j(1)\right)\right]$ and its corresponding MSE under difference choices of Δ among $\Delta \in \{0.02, 0.05, 0.1, 0.2, 0.5, 1, 1.5, 2\}$.

References

[1]. Aalen OO, Stensrud MJ, Didelez V, Daniel R, Røysland K, and Strohmaier S. Time-dependent mediators in survival analysis: Modeling direct and indirect effects with the additive hazards model. Biometrical Journal, 62(3):532–549, 2020. [PubMed: 30779372]

[2]. Akritas MG. Nonparametric survival analysis. Statistical Science, 19(4):615–623, 2004.

[3]. Auranen K, Arjas E, Leino T, and Takala AK. Transmission of pneumococcal carriage in families: A latent Markov process model for binary longitudinal data. Journal of the American Statistical Association, 95(452):1044–1053, 2000.

[4]. Becker NG, Britton T, and O'Neill PD. Estimating vaccine effects on transmission of infection from household outbreak data. Biometrics, 59(3):467–475, 2003. [PubMed: 14601747]

[5]. Becker NG, Britton T, and O'Neill PD. Estimating vaccine effects from studies of outbreaks in household pairs. Statistics in Medicine, 25(6):1079–1093, 2006. [PubMed: 16287206]

[6]. Bhattacharya R, Malinsky D, and Shpitser I. Causal inference under interference and network uncertainty. Uncertainty in Artificial Intelligence, 2019, 2019.

[7]. Cauchemez S, Carrat F, Viboud C, Valleron AJ, and Boëlle P. A Bayesian MCMC approach to study transmission of influenza: application to household longitudinal data. Statistics in Medicine, 23(22):3469–3487, 2004. [PubMed: 15505892]

[8]. Cauchemez S, Temime L, Guillemot D, Varon E, Valleron A-J, Thomas G, and Bo P-Yëlle. Investigating heterogeneity in pneumococcal transmission: a Bayesian MCMC approach applied to a follow-up of schools. Journal of the American Statistical Association, 101 (475):946–958, 2006.

[9]. Cauchemez S, Donnelly CA, Reed C, Ghani AC, Fraser C, Kent CK, Finelli L, and Ferguson NM. Household transmission of 2009 pandemic influenza A (H1N1) virus in the United States. New England Journal of Medicine, 361(27):2619–2627, 2009.

[10]. Chiba Y. A note on bounds for the causal infectiousness effect in vaccine trials. Statistics & Probability Letters, 82(7):1422–1429, 2012.

[11]. Chiba Y. A simple method of measuring vaccine effects on infectiousness and contagion. Open Journal of Statistics, 3(4A):7–15, 2013.

[12]. Chiba Y and Taguri M. Conditional and unconditional infectiousness effects in vaccine trials. Epidemiology, 24(2):336–337, 2013. [PubMed: 23377097]

[13]. Cox D. Planning of Experiments. John Wiley & Sons, New York, 1958.

[14]. Didelez V. Defining causal mediation with a longitudinal mediator and a survival outcome. Lifetime Data Analysis, 25(4):593–610, 2019. [PubMed: 30218418]

[15]. Eck DJ, Morozova O, and Crawford FW. Randomization for the susceptibility effect of an infectious disease intervention in a clustered study population. arXiv preprint arXiv:1808.05593, 2019.

[16]. Fine PEM, Clarkson JA, and Miller E. The Efficacy of Pertussis Vaccines under Conditions of Household Exposure: Further Analysis of the 1978–80 PHLS/ERL Study in 21 Area Health Authorities in England. International Journal of Epidemiology, 17(3):635–642, 9 1988. [PubMed: 3264824]

[17]. Francis T Jr. Evaluation of the 1954 poliomyelitis vaccine field trial: Further studies of results determining the effectiveness of poliomyelitis vaccine (Salk) in preventing paralytic

poliomyelitis. Journal of the American Medical Association, 158(14):1266–1270, 1955. [PubMed: 14392076]

[18]. Golm GT, Elizabeth Halloran M, and Longini IM. Semiparametric methods for multiple exposure mismeasurement and a bivariate outcome in HIV vaccine trials. Biometrics, 55(1):94–101, 1999. [PubMed: 11318183]

[19]. Greenwood M and Yule GU. The statistics of anti-typhoid and anti-cholera inoculations, and the interpretation of such statistics in general. Proceedings of the Royal Society of Medicine, 8(Sect Epidemiol State Med):113, 1915. [PubMed: 19978918]

[20]. Halloran ME and Hudgens MG. Causal inference for vaccine effects on infectiousness. The International Journal of Biostatistics, 8(2), 2012.

[21]. Halloran ME and Hudgens MG. Dependent happenings: a recent methodological review. Current Epidemiology Reports, 3:297–305, 2016. [PubMed: 28133589]

[22]. Halloran ME and Struchiner CJ. Study designs for dependent happenings. Epidemiology, 2(5): 331–338, 1991. [PubMed: 1742381]

[23]. Halloran ME and Struchiner CJ. Causal inference in infectious diseases. Epidemiology, 6 (2): 142–151, 1995. [PubMed: 7742400]

[24]. Halloran ME, Haber M, Longini IM, and Struchiner CJ. Direct and indirect effects in vaccine efficacy and effectiveness. American Journal of Epidemiology, 133(4):323–331, 1991. [PubMed: 1899778]

[25]. Halloran ME, Longini IM, Haber MJ, Struchiner CJ, and Brunet RC. Exposure efficacy and change in contact rates in evaluating prophylactic HIV vaccines in the field. Statistics in Medicine, 13(4):357–377, 1994. [PubMed: 8177986]

[26]. Halloran ME, Struchiner CJ, and Longini IM. Study designs for evaluating different efficacy and effectiveness aspects of vaccines. American Journal of Epidemiology, 146(10):789–803, 1997. [PubMed: 9384199]

[27]. Halloran ME, Longini IM, and Struchiner CJ. Design and interpretation of vaccine field studies. Epidemiological Reviews, 21(1):73–88, 1999.

[28]. Halloran ME, Préziosi MP, and Chu H. Estimating vaccine efficacy from secondary attack rates. Journal of the American Statistical Association, 98(461):38–46, 2003.

[29]. Halloran ME, Longini IM, and Struchiner CJ. Design and Analysis of Vaccine Studies. Springer, 2010.

[30]. Hudgens MG and Halloran ME. Causal vaccine effects on binary postinfection outcomes. Journal of the American Statistical Association, 101(473):51–64, 2006. [PubMed: 19096723]

[31]. Hudgens MG and Halloran ME. Toward causal inference with interference. Journal of the American Statistical Association, 103(482):832–842, 2008. [PubMed: 19081744]

[32]. Kenah E. Contact intervals, survival analysis of epidemic data, and estimation of R_0. Biostatistics, 12(3):548–566, 2010. [PubMed: 21071607]

[33]. Kenah E. Non-parametric survival analysis of infectious disease data. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 75(2):277–303, 2013. [PubMed: 23772180]

[34]. Kenah E. Semiparametric relative-risk regression for infectious disease transmission data. Journal of the American Statistical Association, 110(509):313–325, 2015. [PubMed: 26146425]

[35]. Kenah E, Lipsitch M, and Robins JM. Generation interval contraction and epidemic data analysis. Mathematical Biosciences, 213(1):71–79, 2008. [PubMed: 18394654]

[36]. Lin S-H, Young JG, Logan R, and VanderWeele TJ. Mediation analysis for a survival outcome with time-varying exposures, mediators, and confounders. Statistics in Medicine, 36 (26):4153–4166, 2017. [PubMed: 28809051]

[37]. Longini IM, Koopman JS, Monto AS, and Fox JP. Estimating household and community transmission parameters for influenza. American Journal of Epidemiology, 115(5):736–751, 1982. [PubMed: 7081204]

[38]. Longini IM, Koopman JS, Haber M, and Cotsonis GA. Statistical inference for infectious diseases: risk-specific household and community transmission parameters. American Journal of Epidemiology, 128(4):845–859, 1988. [PubMed: 3421247]

[39]. McSherry GD, Shapiro DE, Coombs RW, McGrath N, Frenkel LM, Britto P, Culnane M, and Sperling RS. The effects of zidovudine in the subset of infants infected with human immunodeficiency virus type-1 (Pediatric AIDS Clinical Trials Group Protocol 076). The Journal of Pediatrics, 134(6):717–724, 1999. [PubMed: 10356140]

[40]. Morozova O, Cohen T, and Crawford FW. Risk ratios for contagious outcomes. Journal of The Royal Society Interface, 15:1–12, 2018.

[41]. Ogburn EL. Challenges to estimating contagion effects from observational data. In Complex Spreading Phenomena in Social Systems, pages 47–64. Springer, 2018.

[42]. Ogburn EL and VanderWeele TJ. Causal diagrams for interference. Statistical science, 29 (4): 559–578, 2014.

[43]. Ogburn EL and VanderWeele TJ. Vaccines, contagion, and social networks. The Annals of Applied Statistics, 11(2):919–948, 2017.

[44]. Ogburn EL, Shpitser I, and Lee Y. Causal inference, social networks, and chain graphs. arXiv preprint arXiv:1812.04990, 2018.

[45]. O'Hagan JJ, Lipsitch M, and Hernán MA. Estimating the per-exposure effect of infectious disease interventions. Epidemiology, 25(1):134, 2014. [PubMed: 24240656]

[46]. O'Neill PD, Balding DJ, Becker NG, Eerola M, and Mollison D. Analyses of infectious disease data from household outbreaks by Markov chain Monte Carlo methods. Journal of the Royal Statistical Society: Series C (Applied Statistics), 49(4):517–542, 2000.

[47]. Orenstein WA, Bernier RH, and Hinman AR. Assessing vaccine efficacy in the field: Further observations. Epidemiologic Reviews, 10(1):212–241, 3 1988. [PubMed: 3066628]

[48]. Pearl J. Causality: models, reasoning and inference. Cambridge University Press, 2000.

[49]. Pearl J. Direct and indirect effects. In Proceedings of the 17th Conference in Uncertainty in Artificial Intelligence, pages 411–420, San Francisco, CA, USA, 2001. Morgan Kaufmann Publishers Inc.

[50]. Rampey AH Jr, Longini IM, Haber M, and Monto AS. A discrete-time model for the statistical analysis of infectious disease incidence data. Biometrics, 48(1):117–128, 1992. [PubMed: 1316178]

[51]. Rhodes PH, Halloran ME, and Longini IM. Counting process models for infectious disease data: Distinguishing exposure to infection from susceptibility. Journal of the Royal Statistical Society: Series B (Methodological), 58(4):751–762, 1996.

[52]. Robins JM and Greenland S. Identifiability and exchangeability for direct and indirect effects. Epidemiology, 3(2):143–155, 1992. [PubMed: 1576220]

[53]. Rosenbaum PR. Interference between units in randomized experiments. Journal of the American Statistical Association, 102(477):191–200, 2007.

[54]. Rubin DB. Causal inference using potential outcomes: Design, modeling, decisions. Journal of the American Statistical Association, 100(469):322–331, 2005.

[55]. Sherman E and Shpitser I. Identification and estimation of causal effects from dependent data. In *Advances in Neural Information Processing Systems* 31, pages 9424–9435. Curran Associates, Inc., 2018.

[56]. Shpitser I. Counterfactual graphical models for longitudinal mediation analysis with unobserved confounding. Cognitive Science, 37(6):1011–1035, 2013. [PubMed: 23899340]

[57]. Shpitser I, Tchetgen Tchetgen E, and Andrews R. Modeling interference via symmetric treatment decomposition. arXiv preprint arXiv:1709.01050, 2017.

[58]. Sjölander A, Frisell T, Kuja-Halkola R, Öberg S, and Zetterqvist J. Carryover effects in sibling comparison designs. Epidemiology, 27(6):852–858, 2016. [PubMed: 27488059]

[59]. Struchiner C and Halloran M. Randomization and baseline transmission in vaccine field trials. Epidemiology & Infection, 135(2):181–194, 2007. [PubMed: 17291359]

[60]. Tchetgen Tchetgen EJ and VanderWeele TJ. On causal inference in the presence of interference. Statistical Methods in Medical Research, 21(1):55–75, 2012. [PubMed: 21068053]

[61]. Tsang TK, Cowling BJ, Fang VJ, Chan K-H, Ip DKM, Leung GM, Peiris JSM, and Cauchemez S. Influenza A virus shedding and infectivity in households. The Journal of Infectious Diseases, 212(9):1420–1428, 2015. [PubMed: 25883385]

[62]. Tsang TK, Fang VJ, Chan K-H, Ip DKM, Leung GM, Peiris JSM, Cowling BJ, and Cauchemez S. Individual correlates of infectivity of influenza A virus infections in households. PLOS ONE, 11(5):1–11, 2016.

[63]. VanderWeele TJ. Direct and indirect effects for neighborhood-based clustered and longitudinal data. Sociological Methods & Research, 38(4):515–544, 2010. [PubMed: 25473138]

[64]. VanderWeele TJ and Tchetgen Tchetgen EJ. Bounding the infectiousness effect in vaccine trials. Epidemiology, 22(5):686–693, 2011. [PubMed: 21753730]

[65]. VanderWeele TJ and Tchetgen Tchetgen EJ. Mediation analysis with time varying exposures and mediators. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 79(3): 917–938, 2017. [PubMed: 28824285]

[66]. VanderWeele TJ, Tchetgen Tchetgen EJ, and Halloran ME. Components of the indirect effect in vaccine trials: identification of contagion and infectiousness effects. Epidemiology, 23 (5):751–761, 2012. [PubMed: 22828661]

[67]. Vansteelandt S, Linder M, Vandenberghe S, Steen J, and Madsen J. Mediation analysis of time-to-event endpoints accounting for repeatedly measured mediators subject to time-varying confounding. Statistics in medicine, 38(24):4828–4840, 2019. [PubMed: 31411779]

[68]. Yang Y, Longini IM, and Halloran ME. Design and evaluation of prophylactic interventions using infectious disease incidence data from close contact groups. Journal of the Royal Statistical Society: Series C (Applied Statistics), 55(3):317–330, 2006.

[69]. Zheng W and van der Laan M. Longitudinal mediation analysis with time-varying mediators and exposures, with application to survival outcomes. Journal of Causal Inference, 5(2):20160006, 2017. [PubMed: 29387520]

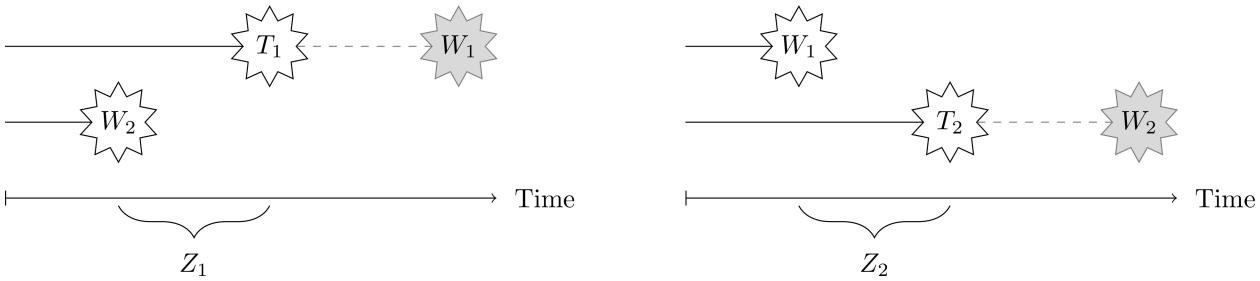

Figure 1:

Illustration of contagion in a two-person partnership. At left, when subject 2 becomes infected first ($W_2 < W_1$), then W_1 is censored, and Z_1 is the remaining time to infection of subject 1. At right, when subject 1 is infected first ($W_1 < W_2$), then W_2 is censored, and Z_2 is the remaining time to infection of subject 2. Informally, the outcome is said to be "contagious" when the distribution of T_i is different from that of W_i.

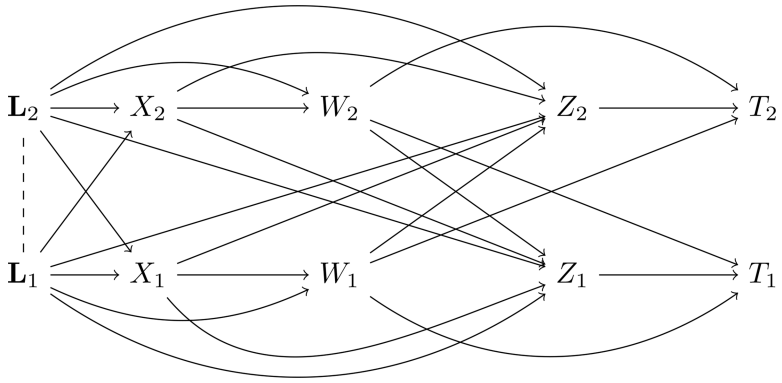

Figure 2:

Causal graphical model for infection outcomes in a two-person partnership, under Assumptions 1–5. Covariates $\mathbf{L}_1$ and $\mathbf{L}_2$ may be dependent within partnerships, and covariates of both subjects may affect the joint treatments (X_1, X_2). The initial infection times W_1 and W_2 are functions of individual covariates and treatments alone by Assumption 1, and thus no arrows exist from X_j to W_i or from L_j to W_i. Subsequent waiting times Z_1 and Z_2 are functions of treatments and covariates of both subjects, and the infection time of the first infected subject. From the decomposition of the infection time (1), the latent additional infection time Z_i and the (possibly latent) time W_i are relevant to exclusive cases of realization of T_i, so they are no arrows between them. The overall infection time T_i is determined by W_i, W_j and Z_i, as specified in (1).

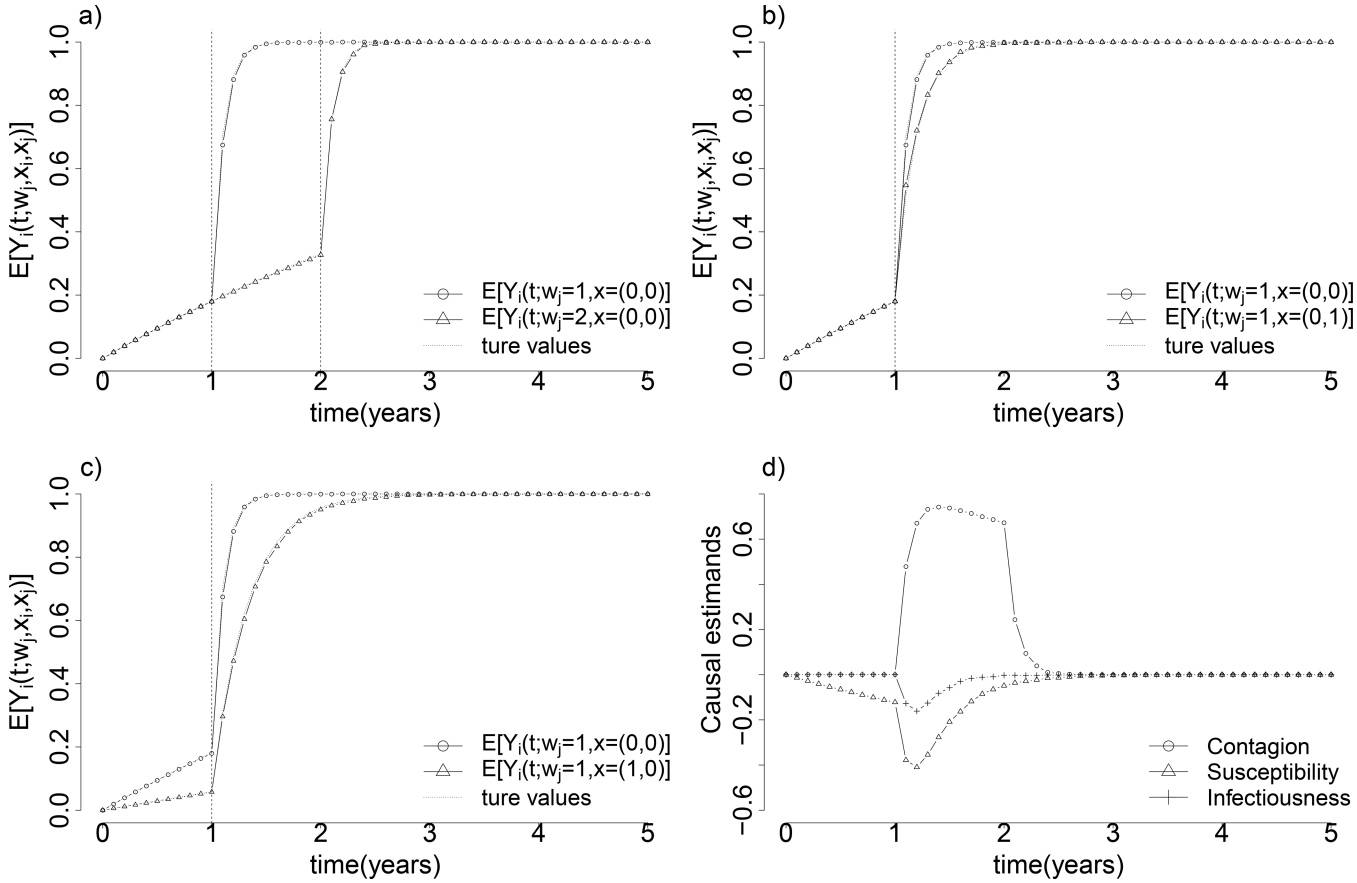

Figure 3:

Illustration of average controlled potential infection outcomes under different values of the infection time w_j and joint treatment **x**, under time-invariant baseline hazards $\alpha(t) = 0.2$ and $\gamma(t - w_j) = 10$ and coefficients $e^{\beta_0} = e^{\beta_1} = 0.3$ and $e^{\sigma} = 0.5$. Contrasts of potential outcomes in (a), (b) and (c) show the controlled contagion effect, the infectiousness effect, and the susceptibility effect evaluated at different times, shown together in (d).

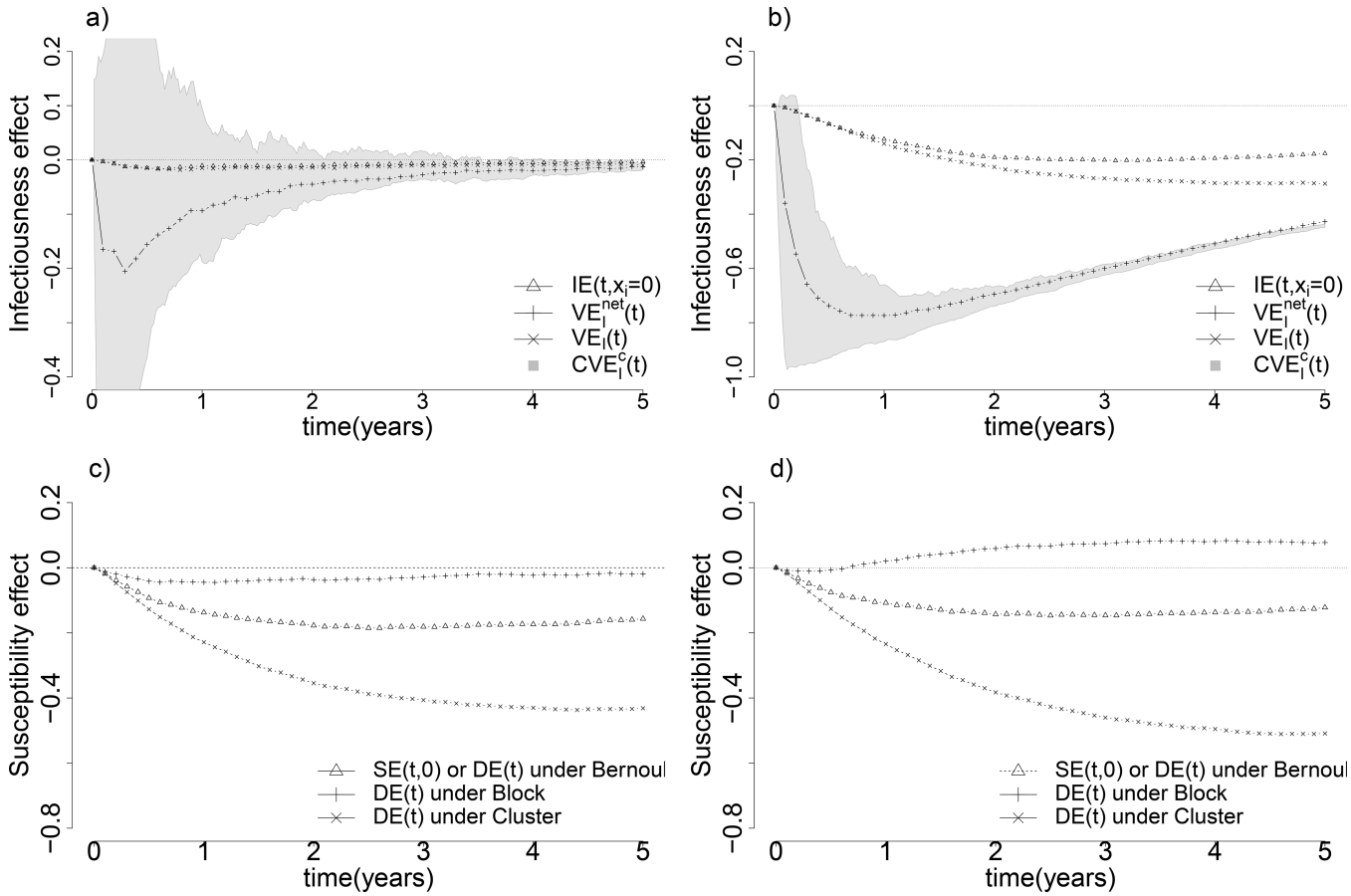

Figure 4:

Comparison of different natural infectiousness and susceptibility effects. Figure a) compares different natural infectiousness effects – natural infectiousness effect IE(t, $x_i = 0$), crude infectiousness effect $\mathrm{VE}_{\mathrm{I}}^{\mathrm{net}}(t)$, the infectiousness defined in mediation analysis $\mathrm{VE}_{\mathrm{I}}(t)$ and bounds identified by principal stratification – when both true susceptibility effect and true infectiousness effect are beneficial ($e^{\beta} = 0.3$, $e^{\sigma} = 0.5$). Similarly, Figure b) shows the same comparison of multiple natural infectiousness effects as in Figure a) when the true infectiousness effect is much stronger than the true susceptibility effect ($e^{\beta} = 0.4$, $e^{\sigma} = 0.01$). Figure c) shows the comparison of different types of natural susceptibility effect – the natural susceptibility effect SE(t, 0), the crude susceptibility effect DE(t) under Bernoulli, Complete, and Cluster randomization – when both true susceptibility effect and true infectiousness effect are beneficial ($e^{\beta} = 0.3$, $e^{\sigma} = 0.5$) as in Figure a). Likewise, Figure d) shows the same comparison of multiple natural susceptibility effects when the true infectiousness effect is much stronger than the true susceptibility effect ($e^{\beta} = 0.4$, $e^{\sigma} = 0.01$). All four graphs are under constant baseline hazards $\alpha(t) = 0.2$ and $\gamma(t) = 10$.

Table 1:

Simulation results showing true values of the natural contagion, susceptibility, infectiousness effects, and alternative estimands defined by Hudgens and Halloran [31], Halloran and Hudgens [20], and VanderWeele et al. [66]. Estimands are evaluated under six different scenarios - (i) constant hazards with $\alpha = 0.2$ and $\gamma = 10$ in (8), (ii) constant hazards without contagion with $\alpha = 0.2$, $\gamma = 0$ in (8), (iii) time-varying hazards with $a = 0.4$, $b = 25$ and $w = 0.5$ in (9), (iv) time-varying external hazard without contagion with $a = 0.4$, $b = 0$ and $w = 0.5$ in (9), (v) time-varying hazards with $a = 0.2$, $b = 40$, $k = 1.5$ and $\theta = 3$ in (10), and (vi) time-varying hazard without contagion with $a = 0.2$, $b = 0$, $k = 1.5$ and $\theta = 3$ in (10), respectively. The effect of vaccination is the same across all scenarios with $e^{\beta_0} = e^{\beta_1} = 0.4$ and $e^{\sigma} = 0.01$. The individual covariates (l_i, l_j) are correlated with $\rho = 0.1$ and coefficients of $e^{\theta_0} = e^{\theta_1} = e^{\theta_2} = 0.95$.

Treatment	CE(*t*, 0)	SE(*t*, 0)	IE(*t*, 0)	DE(*t*)	IDE(*t*)	$\mathrm{VE}_{\mathrm{I}}^{\mathrm{net}}(t)$	$\mathrm{CVE}_{\mathrm{I}}^{c}(t)$
Constant hazards							
Observational	0.12	−0.14	−0.19	−0.16	−0.20	−0.70	-
Bernoulli	0.12	−0.14	−0.19	−0.16	−0.20	−0.70	(−0.73, −0.66)
Block	-	-	-	0.06	-	-	-
Cluster	-	-	-	−0.39	-	-	-
Constant hazards without contagion							
Observational	0.00	−0.18	0.00	−0.18	0.00	−0.01	-
Bernoulli	0.00	−0.18	0.00	−0.18	0.00	−0.01	(−0.25, 0.19)
Block	-	-	-	−0.18	-	-	-
Cluster	-	-	-	−0.18	-	-	-
Time-varying external and decreasing internal hazards							
Observational	0.12	−0.14	−0.20	−0.21	−0.22	−0.51	-
Bernoulli	0.12	−0.14	−0.20	−0.21	−0.22	−0.51	(−0.53, −0.50)
Block	-	-	-	0.08	-	-	-
Cluster	-	-	-	−0.50	-	-	-
Time-varying external and decreasing internal hazards without contagion							
Observational	0.00	−0.28	0.00	−0.28	0.00	−0.02	-
Bernoulli	0.00	−0.28	0.00	−0.28	0.00	−0.02	(−0.43, 0.36)
Block	-	-	-	−0.28	-	-	-
Cluster	-	-	-	−0.28	-	-	-
Time-varying external and increasing-then-decreasing internal hazards							
Observational	0.10	−0.16	−0.17	−0.17	−0.18	−0.64	-
Bernoulli	0.10	−0.16	−0.17	−0.17	−0.18	−0.64	(−0.62, −0.39)
Block	-	-	-	0.02	-	-	-
Cluster	-	-	-	−0.37	-	-	-
Time-varying external and increasing-then-decreasing internal hazards without contagion							
Observational	0.00	−0.18	0.00	−0.18	0.00	−0.01	-
Bernoulli	0.00	−0.18	0.00	−0.18	0.00	−0.01	(−0.43, 0.36)
Block	-	-	-	−0.18	-	-	-

Treatment	CE(*t*, 0)	SE(*t*, 0)	IE(*t*, 0)	DE(*t*)	IDE(*t*)	$\mathrm{VE}_{\mathrm{I}}^{\mathrm{net}}(t)$	$\mathrm{CVE}_{\mathrm{I}}^{c}(t)$
Cluster	-	-	-	−0.18	-	-	-

Table 2:

Simulation results showing true values of the natural contagion, susceptibility, infectiousness effects, and alternative estimands defined by Hudgens and Halloran [31], Halloran and Hudgens [20], and VanderWeele et al. [66]. Estimands are evaluated under six different scenarios - (i) constant hazards with $\alpha = 0.2$ and $\gamma = 10$ in (8), (ii) constant hazards without contagion with $\alpha = 0.2$, $\gamma = 0$ in (8), (iii) time-varying hazards with $a = 0.4$, $b = 25$ and $w = 0.5$ in (9), (iv) time-varying external hazard without contagion with $a = 0.4$, $b = 0$ and $w = 0.5$ in (9), (v) time-varying hazards with $a = 0.2$, $b = 40$, $k = 1.5$ and $\theta = 3$ in (10), and (vi) time-varying hazard without contagion with $a = 0.2$, $b = 0$, $k = 1.5$ and $\theta = 3$ in (10), respectively. The effect of vaccination is the same across all scenarios with $e^{\beta_0} = e^{\beta_1} = 0.4$ and $e^{\sigma} = 0.5$. The individual covariates (l_i, l_j) are correlated with $\rho = 0.1$ and coefficients of $e^{\theta_0} = e^{\theta_1} = e^{\theta_2} = 0.95$.

Treatment	CE(t, 0)	SE(t, 0)	IE(t, 0)	DE(t)	IDE(t)	$\mathrm{VE}_{\mathrm{I}}^{\mathrm{net}}(t)$	$\mathrm{CVE}_{\mathrm{I}}^{c}(t)$
Constant hazards							
Observational	0.14	−0.18	−0.01	−0.20	−0.14	−0.04	-
Bernoulli	0.14	−0.18	−0.01	−0.20	−0.14	−0.04	(−0.08, 0.02)
Block	-	-	-	−0.04	-	-	-
Cluster	-	-	-	−0.36	-	-	-
Constant hazards without contagion							
Observational	0.00	−0.22	0.00	−0.22	0.00	−0.01	-
Bernoulli	0.00	−0.22	0.00	−0.22	0.00	−0.01	(−0.39, 0.19)
Block	-	-	-	−0.22	-	-	-
Cluster	-	-	-	−0.22	-	-	-
Time-varying external and decreasing internal hazards							
Observational	0.15	−0.18	−0.01	−0.23	−0.15	−0.03	-
Bernoulli	0.15	−0.18	−0.01	−0.23	−0.15	−0.03	(−0.04, 0.00)
Block	-	-	-	−0.03	-	-	-
Cluster	-	-	-	−0.44	-	-	-
Time-varying external and increasing-then-decreasing internal hazards without contagion							
Observational	0.00	−0.34	0.00	−0.34	0.00	−0.02	-
Bernoulli	0.00	−0.34	0.00	−0.34	0.00	−0.02	(−0.64, 0.36)
Block	-	-	-	−0.34	-	-	-
Cluster	-	-	-	−0.34	-	-	-
Time-varying external and increasing-then-decreasing internal hazards							
Observational	0.12	−0.21	−0.02	−0.22	−0.13	−0.08	-
Bernoulli	0.12	−0.21	−0.02	−0.22	−0.13	−0.08	(−0.21, 0.07)
Block	-	-	-	−0.08	-	-	-
Cluster	-	-	-	−0.36	-	-	-
Time-varying external and increasing-then-decreasing internal hazards without contagion							
Observational	0.00	−0.22	0.00	−0.22	0.00	−0.01	-
Bernoulli	0.00	−0.22	0.00	−0.22	0.00	−0.01	(−0.64, 0.36)
Block	-	-	-	−0.22	-	-	-

Treatment	**CE(*t*, 0)**	**SE(*t*, 0)**	**IE(*t*, 0)**	**DE(*t*)**	**IDE(*t*)**	$\mathrm{VE}_{\mathrm{I}}^{\mathrm{net}}(t)$	$\mathrm{CVE}_{\mathrm{I}}^{c}(t)$
Cluster	-	-	-	−0.22	-	-	-