Causal Estimands for Infectious Disease Count Outcomes to Investigate the Public Health Impact of Interventions

Authors: Katherine M Jia[1§], Christopher B Boyer[1], Jacco Wallinga[2,3], Marc Lipsitch[1,4]

Affiliations:

[1] Center for Communicable Disease Dynamics, Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA

[2] Centre for Infectious Disease Control, National Institute for Public Health and the Environment, Bilthoven, the Netherlands

[3] Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, the Netherlands

[4] Department of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA

[§] **Correspondence:** Katherine Jia, MSc, Center for Communicable Disease Dynamics, Department of Epidemiology, Harvard T.H. Chan School of Public Health, 677 Huntington Avenue, Boston, MA 02115, USA. +1 (617) 432-7955. (kjia@g.harvard.edu).

ABSTRACT

Causal estimands of infectious disease interventions—direct, indirect, overall, and total effects—are conventionally defined as differences in individual risk under hypothetical treatment conditions. During the coronavirus disease (COVID-19) pandemic, researchers implicitly targeted analogous estimands at the population level by comparing count outcomes (e.g., vaccine-averted deaths) to quantify public health impact of non-pharmaceutical interventions or vaccination campaigns. However, these population-level analogs of conventional estimands have not been rigorously defined. Using potential outcome notation, we introduce a population-level analog of the overall effect and partitioned it into components involving individual-level direct and indirect effects. We further identify conditions under which a population-level analog of direct effect (frequently estimated with empirical data in cases-averted or avertible analyses), can be a useful lower bound of overall effect (arguably the most relevant effect for policy-making and retrospective policy evaluation) at the population level. To illustrate, we describe a susceptible-infected-recovered-death model stratified by vaccination status. When transmission and fatality parameters do not vary and vaccine efficacies do not wane over time, cases averted via direct effect among vaccinated individuals (or cases avertible via direct effect among unvaccinated individuals) is shown to be a lower bound of population-level overall effect—that is, vaccine-averted (or avertible) cases. However, when vaccine efficacies wane, this relation may not hold for avertible cases; when transmission or fatality parameters increase over time, it may not hold for either analysis. By classifying population-level estimands and establishing their relations, this study improves conduct and interpretation of research evaluating impact of infectious disease interventions.

RUNNING HEAD: Effect estimands for infectious disease count outcomes

KEYWORDS: direct effect, overall effect, vaccine-averted deaths, vaccine-preventable deaths

1. INTRODUCTION

The effects of an infectious disease intervention, such as vaccination, are quantified in terms of direct, indirect, overall, and total effects—each defined as a contrast between potential outcomes at the individual level.[1] These estimands consider the individual risk of developing an infectious disease outcome as a function of the individual treatment status and the treatment status of the rest of the group to which the individual belongs.[2] Throughout the coronavirus disease (COVID-19) pandemic, researchers have often targeted the analogs of these estimands at the population level (i.e., estimands defined for count outcomes such as the number of infections, hospitalizations, and deaths in a group of individuals, instead of risk in each individual) to investigate the public health impact of interventions such as vaccination. For example, researchers have estimated the COVID-19 deaths (or infections) averted by vaccination campaigns in the United States,[3,4] Israel,[5,6] Chile,[7] Brazil,[8] and Japan.[9] Similarly, other studies including ours[10] have attempted to estimate vaccine-avertible—that is, how many deaths could have been averted by vaccination, but were not because of a failure to vaccinate a proportion of the population.[10,11] (Note that in our study,[10] large uncertainties exist regarding the exact number of preventable deaths). However, to the best of our knowledge, these population-level estimands have not been rigorously defined in a causal inference framework. Consequently, different studies have measured different kinds of vaccine-averted deaths. Some studies estimated the total deaths averted in a whole country by simulating the epidemic trajectory of an entire country under the counterfactual of no vaccination and compared it with the estimated deaths under the observed vaccination campaign;[3,12] other studies estimated vaccine-averted deaths via the direct effect among the vaccinated individuals by taking the difference in death rates between the vaccinated and unvaccinated groups and multiplying the difference with the number of vaccinated individuals with no previous documented infections.[6–8]

Although both types of studies aimed to quantify "vaccine-averted deaths," they targeted different estimands, but the existing conceptual framework was not designed to distinguish between these different population-level estimands.

This study defines and classifies the population-level analogs of causal estimands for infectious disease interventions using potential outcome notation, such that researchers can use and distinguish the estimands to evaluate public health impact of infectious disease interventions. Based on an infectious disease transmission model, we identify the conditions under which direct effect (easiest and most frequently estimated among all types of effects) is a lower bound of overall effect (arguably the most relevant effect for policy-making and retrospective policy evaluation) at the population level.[13,14]

2. INDIVIDUAL-LEVEL DIRECT, INDIRECT, OVERALL, AND TOTAL EFFECTS

Hudgens and Halloran[2] previously defined causal estimands for individual-level direct, indirect, overall, and total effects. We follow their setup and notation but add an additional index t to outcomes to represent cumulative incidence up to time t. Consider a two-stage randomized experiment in which there are m groups indexed by $i = 1, \dots, m$, such that each group consists of N individuals indexed by $j = 1, \dots, N$. Partial interference is assumed: Individuals make contacts (leading to interference) within the same group, but individuals in different groups make no contacts. For ease of exposition, assume interest lies in quantifying the effect of vaccination which is a one-time event before the start of an outbreak. Let $A_{ij} = 1$ if individual j in each group i is vaccinated and $A_{ij} = 0$ otherwise. Let $\boldsymbol{A_i} = (A_{i1}, A_{i2}, \dots, A_{iN})$ and $\boldsymbol{A_{i,-j}} = (A_{i1}, A_{i2}, \dots, A_{ij-1}, A_{ij+1}, \dots, A_{iN})$, hereafter referred to as allocation programs.[14] Let $\mathbf{a_i}$ and $\mathbf{a_{i,-j}}$ denote possible realizations of $\boldsymbol{A_i}$ and $\boldsymbol{A_{i,-j}}$, respectively. Let $\mathcal{A}(N)$ denote the set of all possible 2^N vaccine allocations for a group of size N, for which $\mathbf{a_i} \in \mathcal{A}(N)$. Let $Y_{ij}(t, \mathbf{a_i})$ denote the

potential binary outcome for individual j in group i at time t with allocation program $\mathbf{a}_{\boldsymbol{i}}$ and let $Y_{ij}(t, \mathbf{a}_{\boldsymbol{i},-\boldsymbol{j}}, a)$ denote the potential binary outcome at time t when individual j has vaccination status a and the rest of group i has vaccination status $\mathbf{a}_{\boldsymbol{i},-\boldsymbol{j}}$. Hudgens and Halloran[2] define the *marginal individual average potential outcome* as

$$\bar{Y}_{ij}(t;\alpha) \equiv \sum_{\mathbf{s}\in\mathcal{A}(N)} Y_{ij}(t, \mathbf{a}_i = \mathbf{s}) \Pr_{\alpha}(\boldsymbol{A}_{\boldsymbol{i}} = \mathbf{s}),$$

and the *individual average potential outcome*[2] as:

$$\bar{Y}_{ij}(t, a; \alpha) \equiv \textstyle\sum_{\boldsymbol{s}\in\mathcal{A}(N-1)} Y_{ij}\big(t, \mathbf{a}_{\boldsymbol{i},-\boldsymbol{j}} = \boldsymbol{s}, a_{ij} = a\big) \Pr_{\alpha}\big(\boldsymbol{A}_{\boldsymbol{i},-\boldsymbol{j}} = \boldsymbol{s} | A_{ij} = a\big).$$

where $\Pr_{\alpha}(\cdot)$ is the probability distribution of vaccine allocation program $\boldsymbol{A}_{\boldsymbol{i}}$ with parameter $\alpha \in [0,1]$ representing the proportion vaccinated within group i (Note here we use type A parameterization, which gives same effect definitions as type B parameterization suggested by VanderWeele and Tchetgen Tchetgen[14] when N is large). Hudgens and Halloran[2] further define group average potential outcomes $\bar{Y}_i(t;\alpha) \equiv \sum_{j=1}^{N} \bar{Y}_{ij}(t;\alpha)\,/N$ and $\bar{Y}_i(t, a; \alpha) \equiv \sum_{j=1}^{N} \bar{Y}_{ij}(t, a; \alpha)/N$ by averaging over individuals within groups. Assume that the m groups are independently drawn from an infinite superpopulation of groups and let $\pi_{\alpha}(t) = E_{\alpha}\{\bar{Y}_i(t;\alpha)\}$ denote the probability of having developed the outcome by time t under the counterfactual setting where the proportion vaccinated is α in the superpopulation of groups.[15] Similarly, let $\pi_{a,\alpha}(t) = E_{\alpha}\{\bar{Y}_i(t, a; \alpha)\}$ denote the probability of having developed the outcome by time t when an individual has vaccination status a and the group has a fixed proportion α of vaccinated individuals. We now define population average individual direct, indirect, total, and overall effects[2] in the superpopulation (Note despite "population average" nomenclature, which is retain from Hudgens and Halloran[2], in our setup these are still individual level effects because they are risks for typical individuals in the superpopulation rather than counts).

Individual-level Direct Effect (IDE)

$$IDE(t, \alpha) = \pi_{0,\alpha}(t) - \pi_{1,\alpha}(t) \quad (1)$$

The individual-level direct effect[2] compares the probability of having developed the outcome by time t when an individual is unvaccinated versus when vaccinated, holding fixed the proportion vaccinated (α).

Individual-level Indirect Effects (IIEs)

$$IIE^{Unvax}(t, \alpha_0, \alpha_1) = \pi_{0,\alpha_0}(t) - \pi_{0,\alpha_1}(t) \quad (2)$$

The individual-level indirect effect[2] compares the probability of having developed the outcome by time t for an unvaccinated individual in a group with α_0 proportion vaccinated versus with α_1 proportion vaccinated. This effect, defined by Hudgens and Halloran[2], is hereafter referred to as the individual-level indirect effect for the unvaccinated. Indirect effect can be analogously defined when an individual is vaccinated as follows:[15]

$$IIE^{Vax}(t, \alpha_0, \alpha_1) = \pi_{1,\alpha_0}(t) - \pi_{1,\alpha_1}(t) \quad (3)$$

which is hereafter referred to as the individual-level indirect effect for the vaccinated.

Individual-level Overall Effect (IOE)

$$IOE(t, \alpha_0, \alpha_1) = \pi_{\alpha_0}(t) - \pi_{\alpha_1}(t) \quad (4)$$

The individual-level overall effect[2] compares the probability of having developed the outcome by time t for a typical individual in a group with α_0 proportion vaccinated versus with α_1 proportion vaccinated.

Individual-level Total Effect (ITE)

$$ITE(t, \alpha_0, \alpha_1) = \pi_{0,\alpha_0}(t) - \pi_{1,\alpha_1}(t) \quad (5)$$

The individual-level total effect[2] compares the probability of having developed the outcome by time t when an individual is unvaccinated in a group with α_0 proportion vaccinated versus

when the individual is vaccinated in a group with α_1 proportion vaccinated. $ITE(t, \alpha_0, \alpha_1)$ can be decomposed into $IDE(t, \alpha_1)$ and $IIE^{Unvax}(t, \alpha_0, \alpha_1)$ by adding and subtracting the term $\pi_{0,\alpha_1}(t)$:

$$ITE(t, \alpha_0, \alpha_1) = \pi_{0,\alpha_1}(t) - \pi_{1,\alpha_1}(t) + \pi_{0,\alpha_0}(t) - \pi_{0,\alpha_1}(t)$$
$$= IDE(t, \alpha_1) + IIE^{Unvax}(t, \alpha_0, \alpha_1). \quad (6.1)$$

Alternatively, $ITE(t, \alpha_0, \alpha_1)$ can also be decomposed into $IDE(t, \alpha_0)$ and $IIE^{Vax}(t, \alpha_0, \alpha_1)$ by adding and subtracting the term $\pi_{1,\alpha_0}(t)$:

$$ITE(t, \alpha_0, \alpha_1) = \pi_{0,\alpha_0}(t) - \pi_{1,\alpha_0}(t) + \pi_{1,\alpha_0}(t) - \pi_{1,\alpha_1}(t)$$
$$= IDE(t, \alpha_0) + IIE^{Vax}(t, \alpha_0, \alpha_1). \quad (6.2)$$

Individual-level Overall Effect Partitioning

Theorem 1 (individual-level overall effect partitioning).

$$IOE(t, \alpha_0, \alpha_1) = \alpha_0 \cdot IIE^{Vax}(t, \alpha_0, \alpha_1) + (\alpha_1 - \alpha_0) \cdot ITE(t, \alpha_0, \alpha_1) + (1 - \alpha_1) \cdot IIE^{Unvax}(t, \alpha_0, \alpha_1).$$

Theorem 1 is proved in eAppendix 1 and graphically illustrated in Figure 1. Theorem 1 expresses $IOE(\alpha_0, \alpha_1)$ as a weighted average of three effects: 1) $IIE^{Vax}(t, \alpha_0, \alpha_1)$, 2) $ITE(t, \alpha_0, \alpha_1)$, and 3) $IIE^{Unvax}(t, \alpha_0, \alpha_1)$. Intuitively, if individuals are classified by their vaccination status under a pair of counterfactuals wherein the group has α_0 or α_1 proportion vaccinated ($\alpha_1 > \alpha_0$), then $IIE^{Vax}(t, \alpha_0, \alpha_1)$ is in operation for proportion α_0 of individuals who are vaccinated under both counterfactuals; $ITE(t, \alpha_0, \alpha_1)$ is in operation for proportion $\alpha_1 - \alpha_0$ of individuals who are vaccinated under α_1 but unvaccinated under α_0; and $IIE^{Unvax}(t, \alpha_0, \alpha_1)$ is in operation for proportion $1 - \alpha_1$ of individuals who are unvaccinated under both counterfactuals. Previously, Hudgens and Halloran[2] showed that when $\alpha_0 = 0$, $IOE(t, \alpha_0, \alpha_1)$ is the weighted sum $\alpha_1 \cdot ITE(t, \alpha_0, \alpha_1) + (1 - \alpha_1) \cdot IIE^{Unvax}(t, \alpha_0, \alpha_1)$. In this study, the partitioning of $IOE(t, \alpha_0, \alpha_1)$ has been generalized to any α_0 and α_1 as in Theorem 1.

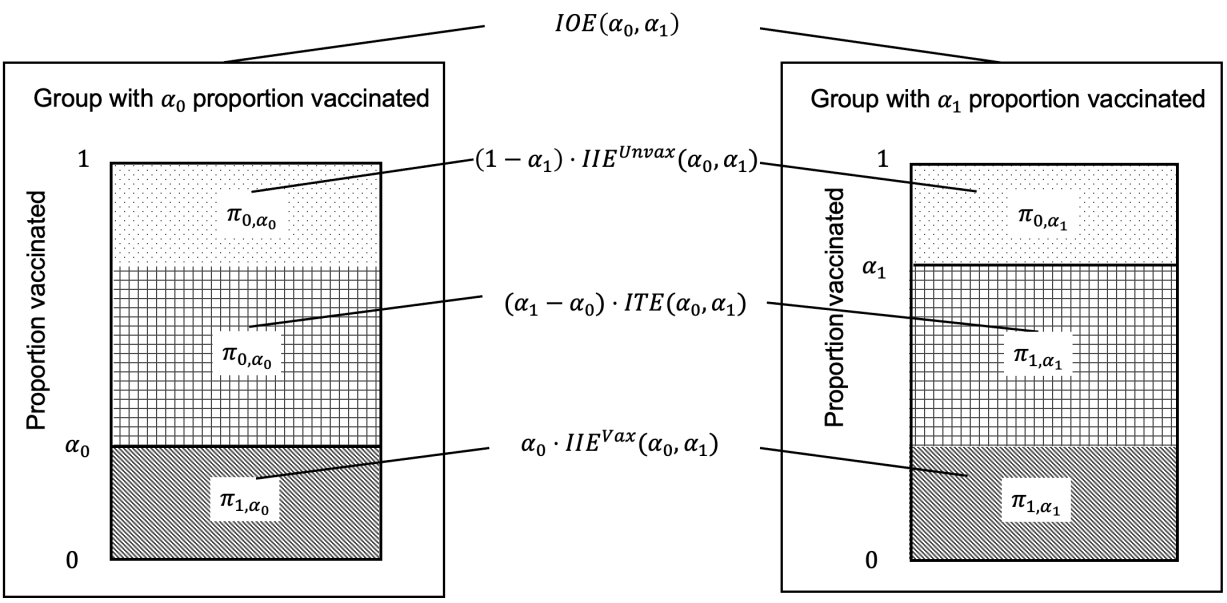

Figure 1. Graphical illustration on partitioning individual-level overall effect. Time notation (t) is suppressed for the ease of exposition. The two rectangles represent a pair of counterfactuals wherein the group has α_0 or α_1 proportion vaccinated ($\alpha_1 > \alpha_0$). Individuals fall into three categories based on their vaccination status under the counterfactuals: 1) the dotted region represents those who are unvaccinated under both counterfactuals (hereafter referred to as "never-vaccinated") and for whom $IIE^{Unvax}(\alpha_0, \alpha_1)$ is in operation; 2) the gridded region represents those who are unvaccinated under α_0 but vaccinated under α_1 and for whom $ITE(\alpha_0, \alpha_1)$ is in operation (hereafter referred to as "additionally-vaccinated"); and 3) the stripped region represents those who are vaccinated under both counterfactuals (hereafter referred to as "always-vaccinated") and for whom $IIE^{Vax}(\alpha_0, \alpha_1)$ is in operation. Theorem 1 shows that $IOE(\alpha_0, \alpha_1)$ is a weighted average of three effects: 1) $IIE^{Unvax}(\alpha_0, \alpha_1)$, 2) $ITE(\alpha_0, \alpha_1)$, and 3) $IIE^{Vax}(\alpha_0, \alpha_1)$, each weighted by the proportion of individuals for whom the effect is in operation respectively: 1) $1 - \alpha_1$ for the never-vaccinated, 2) $\alpha_1 - \alpha_0$ for the additionally-vaccinated, and 3) α_0 for the always-vaccinated.

3. POPULATION-LEVEL DIRECT AND OVERALL EFFECTS

Motivating Examples and Causal Questions

During the COVID-19 pandemic, determining the total number of infections (or deaths) averted by vaccination has been of great public health interest.[3–9,12,16,17] Vaccine-averted infections (or deaths) is a population-level estimand based on the causal question: How many infections (or deaths) have been averted under the current proportion vaccinated (α_1) compared to the counterfactual in the absence of vaccination ($\alpha_0 = 0$)?

Alternatively, researchers have also shown interests in estimating the vaccine-avertible deaths, that is those that could have been averted by vaccination but were not because of a failure to vaccinate.[10,11] The causal question in this case is: How many infections (or deaths) could have been averted under full vaccination ($\alpha_2 = 1$), but were not averted given the current proportion vaccinated (α_1)?

In general, population-level estimands can be defined by comparing the number of infections (or deaths) in a typical group under the observed proportion vaccinated (α_1) versus when the group has a lower (denoted as α_0, where $\alpha_0 < \alpha_1$) or higher proportion (denoted as α_2, where $\alpha_1 < \alpha_2$) vaccinated.[11] We term these as "population-level" estimands because previous literature has referred to vaccine-averted infections (or deaths) as a population(-level) impact[9,16] or an estimand with population-level implications.[5] Note that these estimands compare count outcomes that could be observed at the group (or cluster) level of a two-stage randomized experiment.[2]

Population-level Overall Effects

The population-level overall effect (POE) directly answers the two aforementioned causal questions on vaccine-averted and vaccine-avertible cases. Therefore, effects of counterfactual

vaccination programs such as the vaccine averted and avertible cases above, are best captured by POE.

For $\alpha_1 > \alpha_0$, POE is defined as:

$$POE(t, \alpha_0, \alpha_1) = N \cdot IOE(t, \alpha_0, \alpha_1). \quad (7.1)$$

For $\alpha_2 > \alpha_1$, POE is defined as:

$$POE(t, \alpha_1, \alpha_2) = N \cdot IOE(t, \alpha_1, \alpha_2). \quad (7.2)$$

Note that mathematical modelling studies implicitly refer to POE when estimating vaccine-averted deaths by simulating the epidemic trajectory under the counterfactual with a hypothetical proportion vaccinated (e.g., $\alpha_0 = 0$) and comparing it with the estimated number under the current vaccination campaign (α_1).[3,4]

Moreover, by Theorem 1 and substituting equation (6.1) into $ITE(t, \alpha_0, \alpha_1)$, equation (7.1) can be decomposed into:

$$\begin{aligned} POE(t, \alpha_0, \alpha_1) &= N \cdot IOE(t, \alpha_0, \alpha_1) \\ &= N \cdot \alpha_0 \cdot IIE^{Vax}(t, \alpha_0, \alpha_1) + N \cdot (\alpha_1 - \alpha_0) \cdot \left(IDE(t, \alpha_1) + IIE^{Unvax}(t, \alpha_0, \alpha_1)\right) \\ &+ N \cdot (1 - \alpha_1) \cdot IIE^{Unvax}(t, \alpha_0, \alpha_1) \\ &= N \cdot \alpha_0 \cdot IIE^{Vax}(t, \alpha_0, \alpha_1) + N \cdot (\alpha_1 - \alpha_0) \cdot IDE(t, \alpha_1) + N \cdot (1 - \alpha_0) \\ &\cdot IIE^{Unvax}(t, \alpha_0, \alpha_1) \quad (8.1) \end{aligned}$$

for $\alpha_1 > \alpha_0$. The first term on the right-hand side of the last line of equation (8.1) scales $IIE^{Vax}(t, \alpha_0, \alpha_1)$ by number vaccinated under α_0. The second term scales $IDE(t, \alpha_1)$ by the additional number vaccinated under α_1 compared to α_0. The third term scales $IIE^{Unvax}(t, \alpha_0, \alpha_1)$ by number unvaccinated under α_0. Similarly, by decomposing equation (7.2) based on Theorem 1 and then substituting equation (6.2) into $ITE(t, \alpha_1, \alpha_2)$, we have:

$$POE(t, \alpha_1, \alpha_2) = N \cdot IOE(t, \alpha_1, \alpha_2)$$
$$= N \cdot \alpha_2 \cdot IIE^{Vax}(t, \alpha_1, \alpha_2) + N \cdot (\alpha_2 - \alpha_1) \cdot IDE(t, \alpha_1) + N \cdot (1 - \alpha_2) \cdot IIE^{Unvax}(t, \alpha_1, \alpha_2) \quad (8.2)$$

for $\alpha_2 > \alpha_1$. The first term on the right-hand side of equation (8.2) scales $IIE^{Vax}(t, \alpha_0, \alpha_1)$ by number vaccinated under α_2. The second term scales $IDE(t, \alpha_1)$ by the additional number vaccinated under α_2 compared to α_1. The third term scales $IIE^{Unvax}(t, \alpha_0, \alpha_1)$ by number unvaccinated under α_2.

Population-level Direct Effects

Most empirical studies only have data to estimate $IDE(t, \alpha_1)$ based on a single group of individuals under the current proportion vaccinated (α_1), but no data from other group(s) under a lower (α_0) or higher proportion vaccinated (α_2); therefore, most empirical studies cannot estimate $POE(t, \alpha_0, \alpha_1)$ or $POE(t, \alpha_1, \alpha_2)$, which are of interest to policy-making and retrospective policy evaluation.[13]

To generalize the individual-level direct effect for count outcomes, we look for a quantity that can be estimated using commonly available data and that is informative on population-level overall effect. A good candidate is $IDE(t, \alpha_1)$ that most empirical studies can estimate. Some studies have estimated deaths averted via direct effect among vaccinated individuals[6–9] based on formulas similar to $N \cdot (\alpha_1 - \alpha_0) \cdot IDE(t, \alpha_1)$ by setting $\alpha_0 = 0$, while other studies have estimated deaths avertible via direct effect among unvaccinated individuals[10,11] based on formulas similar to $N \cdot (\alpha_2 - \alpha_1) \cdot IDE(t, \alpha_1)$ by setting α_2 to be greater than α_1 (Note these studies have also considered increases in proportion vaccinated over time, such as occurred during a vaccination campaign, which here we ignore for simplicity).

Following the literature,[6–11,16] let the population-level direct effect (PDE) for any α, $\alpha' \in [0,1]$ be:

$$PDE(t, \alpha, \alpha') = N \cdot |\alpha' - \alpha| \cdot IDE(t, \alpha').$$

In particular, for $\alpha_1 > \alpha_0$, we have

$$PDE(t, \alpha_0, \alpha_1) = N \cdot (\alpha_1 - \alpha_0) \cdot IDE(t, \alpha_1) \quad (9.1)$$

and for $\alpha_2 > \alpha_1$,

$$PDE(t, \alpha_2, \alpha_1) = N \cdot (\alpha_2 - \alpha_1) \cdot IDE(t, \alpha_1). \quad (9.2)$$

Note that $PDE(t, \alpha_0, \alpha_1)$ or $PDE(t, \alpha_2, \alpha_1)$ is not a meaningful causal effect by itself because $IDE(t, \alpha_1)$ is conditional on α_1 only and does not account for the change in individual-level direct effect under the counterfactual when the proportion vaccinated is α_0 or α_2 instead of α_1. However, PDE can be a useful lower bound of POE because now equations (8.1) and (8.2) can be written as:

$$POE(t, \alpha_0, \alpha_1) = N \cdot \alpha_0 \cdot IIE^{Vax}(t, \alpha_0, \alpha_1) + PDE(t, \alpha_0, \alpha_1) + N \cdot (1 - \alpha_0) \cdot IIE^{Unvax}(t, \alpha_0, \alpha_1) \quad (10.1)$$

for $\alpha_1 > \alpha_0$, and

$$POE(t, \alpha_1, \alpha_2) = N \cdot \alpha_2 \cdot IIE^{Vax}(t, \alpha_1, \alpha_2) + PDE(t, \alpha_2, \alpha_1) + N \cdot (1 - \alpha_2) \cdot IIE^{Unvax}(t, \alpha_1, \alpha_2) \quad (10.2)$$

for $\alpha_2 > \alpha_1$.

Consider Claim 1:

$$POE(t, \alpha_0, \alpha_1) \geq PDE(t, \alpha_0, \alpha_1) \quad \forall\, \alpha_1 > \alpha_0,$$

and

$$POE(t, \alpha_1, \alpha_2) \geq PDE(t, \alpha_2, \alpha_1) \quad \forall\, \alpha_2 > \alpha_1.$$

In words, Claim 1 asserts that PDE is a lower bound of POE. If Claim 1 is true, PDE, which can be estimated using commonly available data,[6–11,16] is a useful lower bound of POE that is relevant for policy-making and retrospective policy evaluation requiring samples from a population of groups as in group-randomized studies.[2] Also consider two special cases of Claim 1.

Claim 1a (vaccine-averted cases):

$$POE(t, 0, \alpha_1) \geq PDE(t, 0, \alpha_1)$$

In words, Claim 1a asserts that for $\alpha_1 > \alpha_0 = 0$, cases averted via direct effect of current vaccination among the vaccinated individuals $PDE(t, 0, \alpha_1)$ is a lower bound of total vaccine-averted cases among both vaccinated and unvaccinated individuals $POE(t, 0, \alpha_1)$.

Claim 1b (vaccine-avertible cases):

$$POE(t, \alpha_1, 0.9) \geq PDE(t, 0.9, \alpha_1)$$

In words, Claim 1b asserts that for $\alpha_1 < \alpha_2 = 0.9$, cases avertible via direct effect of current vaccination among some unvaccinated individuals $PDE(t, 0.9, \alpha_1)$ is a lower bound of vaccine-avertible cases among both vaccinated and unvaccinated individuals $POE(t, \alpha_1, 0.9)$ comparing current proportion vaccinated (α_1) to a proportion of $\alpha_2 = 0.9$. Note that we do not compare to full vaccination ($\alpha_2 = 1$) because full vaccination is not realistic when some individuals cannot be vaccinated due to contraindications. In the following sections, we use a susceptible-infected-recovered-death (SIRD) model to verify the conditions under which the Claims may or may not hold.

Based on the effect partitioning results in 10.1 or 10.2, Claims 1a and 1b would be true if IIE^{Vax} and IIE^{Unvax} are non-negative. However, it is not immediately intuitive when we might

expect that to be the case. Therefore, in the next two sections we describe a transmission model that we use to check Claims 1a and 1b under various scenarios (Section 4), and then describe conditions under which the population direct effect is and is not a lower bound for the population overall effect, using both analytical and simulation-based approaches (Section 5).

4. TRANSMISSION MODEL

The SIRD Model with Vaccination at Baseline

To model the impact of vaccination on infection and death, a SIRD model is used to represent a well-mixed group in a two-stage randomized experiment assuming partial interference.[2] The model consists of four states for a vaccinated or unvaccinated individual—susceptible, infectious, recovered, and death due to infection. We assume that the group has been randomly assigned with a vaccination policy wherein the proportion vaccinated is α, and individuals have been randomly assigned with vaccination status a at baseline ($a = 1$ for vaccinated and $a = 0$ for unvaccinated; for equation [11], subscript v denotes vaccinated and u for unvaccinated). The vaccine is "leaky" in protecting against infection and infection-related death—that is, vaccinated individuals have the susceptibility reduced by a factor θ against infection (i.e., vaccine efficacy against infection $[VE_{infection}]$ is $(1-\theta)\%$) and have the susceptibility reduced by another factor κ against death (i.e., vaccine efficacy against death given infection $[VE_{death|infection}]$ is $(1-\kappa)\%$). Individuals mix homogeneously, such that each vaccinated or unvaccinated susceptible individual is equally likely to be infected by any infectious individual. The transmission dynamics are:

$$\left.\begin{array}{c} dS_{u,\alpha}(t)/dt = -\lambda_\alpha(t) \cdot S_{u,\alpha}(t) \\ dS_{v,\alpha}(t)/dt = -\theta \cdot \lambda_\alpha(t) \cdot S_{v,\alpha}(t) \\ dI_{u,\alpha}(t)/dt = \lambda_\alpha(t) \cdot S_{u,\alpha}(t) - \gamma \cdot I_{u,\alpha}(t) \\ dI_{v,\alpha}(t)/dt = \theta \cdot \lambda_\alpha(t) \cdot S_{v,\alpha}(t) - \gamma \cdot I_{v,\alpha}(t) \\ dR_{u,\alpha}(t)/dt = (1-\mu) \cdot \gamma \cdot I_{u,\alpha}(t) \\ dR_{v,\alpha}(t)/dt = (1-\kappa \cdot \mu) \cdot \gamma \cdot I_{v,\alpha}(t) \\ dD_{u,\alpha}(t)/dt = \mu \cdot \gamma \cdot I_{u,\alpha}(t) \\ dD_{v,\alpha}(t)/dt = \kappa \cdot \mu \cdot \gamma \cdot I_{v,\alpha}(t) \end{array}\right\} (11)$$

where $\gamma =$ recovery rate, $\lambda_\alpha(t) = \beta \cdot \frac{I_{u,\alpha}(t)+I_{v,\alpha}(t)}{N(t)}$ is the hazard rate of infection, with β the number of effective contacts made by a typical infectious individual per unit time, and $\mu =$ probability of death due to infection. In equation (11), $S_{u,\alpha}(t)$ and $S_{v,\alpha}(t)$ denote, respectively, the number of susceptible individuals who are unvaccinated and vaccinated, $I_{u,\alpha}(t)$ and $I_{v,\alpha}(t)$ for the infectious individuals, $R_{u,\alpha}(t)$ and $R_{v,\alpha}(t)$ for the recovered individuals who are no longer at risk, and $D_{u,\alpha}(t)$ and $D_{v,\alpha}(t)$ for those who died due to infection. $N(t)$ denotes the sum of all compartments at time t. eFigure 1 shows the model flowchart, and eTable 1 shows the parameter values used in simulation.

Software

All model simulations and visualization are conducted using R 4.2.2 (R Foundation for Statistical Computing, Vienna, Austria).[18] All models are implemented using R package *odin*.[19] Code is available at https://github.com/katjia/population_level_effects.

5. WHEN IS POPULATION-LEVEL DIRECT EFFECT A LOWER BOUND OF POPULATION-LEVEL OVERALL EFFECT?

Scenarios

We check the Claims under 5 scenarios (Table). Scenario 1 refers to the SIRD model in equation (11) with time-invariant parameters. eTable 1 lists the model parameters for which one or two parameter(s) varies under each Scenario separately: Scenario 2 increases the number of

effective contacts made by a typical infectious individual per day (β) from 0.15 to 0.6 from Day 300 onwards; Scenario 3 increases probability of death due to infection (μ) from 0.01 to 0.1 from Day 300 onwards; Scenario 4 allows both $VE_{infection}$ and $VE_{death|infection}$ to wane linearly after Day 100 reaching 0% at Day 300; and Scenario 5 combines Scenarios 2 (increasing β) and 4 (waning VEs).

TABLE. Scenarios under which the Claims may or may not hold [a]

Scenario	**Parameters**	**Claim 1**[b]	**Claim 1a**[b]	**Claim 1b**[b]
Scenario 1	Time-invariant parameters	+	+	+
Scenario 2	β increases from 0.15 to 0.6 from Day 300 onwards	-	-	-
Scenario 3	μ increases from 0.01 to 0.1 from Day 300 onwards	-	-	-
Scenario 4	$VE_{infection}$ and $VE_{death|infection}$ wane linearly after Day 100, reaching 0% at Day 300	-	+	-
Scenario 5	Scenarios 2 and 4 combined	-	-	-

β is the number of effective contacts made by a typical infectious individual per day; μ is the probability of death due to infection; $VE_{infection}$ is the vaccine efficacy against infection; $VE_{death|infection}$ is the vaccine efficacy against death given infection.

[a] Positive sign (+) indicates that the Claim holds; negative sign (-) otherwise.

[b] Claim 1 holds only if Claims 1a and 1b hold.

Proof of Claim 1 at the end of outbreak under Scenario 1

For Scenario 1 (i.e., time-invariant parameters), we prove Claim 1 in eAppendix 3—that is, population-level direct effects are lower bounds of overall effects for any two proportions vaccinated—holds in the SIRD model at the end of outbreak (i.e., at $t \to \infty$).

Simulations

For $t < \infty$, we first verify Claims 1a and 1b (special cases of Claim 1) through simulation based on parameters specified in eTable 1 and initial conditions in eTable 2. We specify $\alpha_0 = 0$ versus $\alpha_1 = 0.7$ in the pair of trajectories to verify Claim 1a and $\alpha_1 = 0.7$ versus $\alpha_2 = 0.9$ to verify Claim 1b. If Claims 1a and 1b both hold for the specified parameters, Latin hypercube sampling is conducted to generate alternative sets of proportions vaccinated and model parameters to verify the full Claim 1. If only one of Claim 1a or 1b holds, Latin hypercube sampling is conducted to verify that the Claim holds under alternative proportion vaccinated and model parameters (e.g., trying alternative values for α_1 while fixing $\alpha_0 = 0$ for Claim 1a).

Briefly, Claim 1 only holds under Scenario 1 (i.e., time-invariant parameters), but it does not hold under any other Scenarios. The Table summarizes the results. Figures 2 and 3 show trajectories of population-level effects throughout the epidemic to verify Claims 1a and 1b, respectively. Figures 2 and 3 show that Claims 1a and 1b hold under Scenario 1. Moreover, given alternative sets of proportions vaccinated and model parameters, Latin hypercube sampling verifies that Claim 1 holds under Scenario 1 (eAppendix 4). Under Scenario 2 where β increases, Claims 1a and 1b do not hold: population-level direct effects are not lower bounds of overall effects (Figures 2 and 3) due to negative indirect effects (eFigures 3 to 5). Under Scenario 3 where μ increases, Claims 1a and 1b hold for infections but not deaths due to negative indirect effects for death (eFigures 3 to 5). Under Scenario 4 where VEs wane, only Claim 1a (vaccine-averted cases) holds (Figure 2). Given different values for α_1 (while holding constant $\alpha_0 = 0$) and alternative sets of model parameters, Latin hypercube sampling verifies that Claim 1a holds under Scenario 4 (eAppendix 4). However, Claim 1b (vaccine-avertible cases) does not hold (Figure 3) due to

negative indirect effects (eFigures 4 and 5). Finally, under Scenario 5 where β increases and VEs wane, Claims 1a and 1b do not hold.

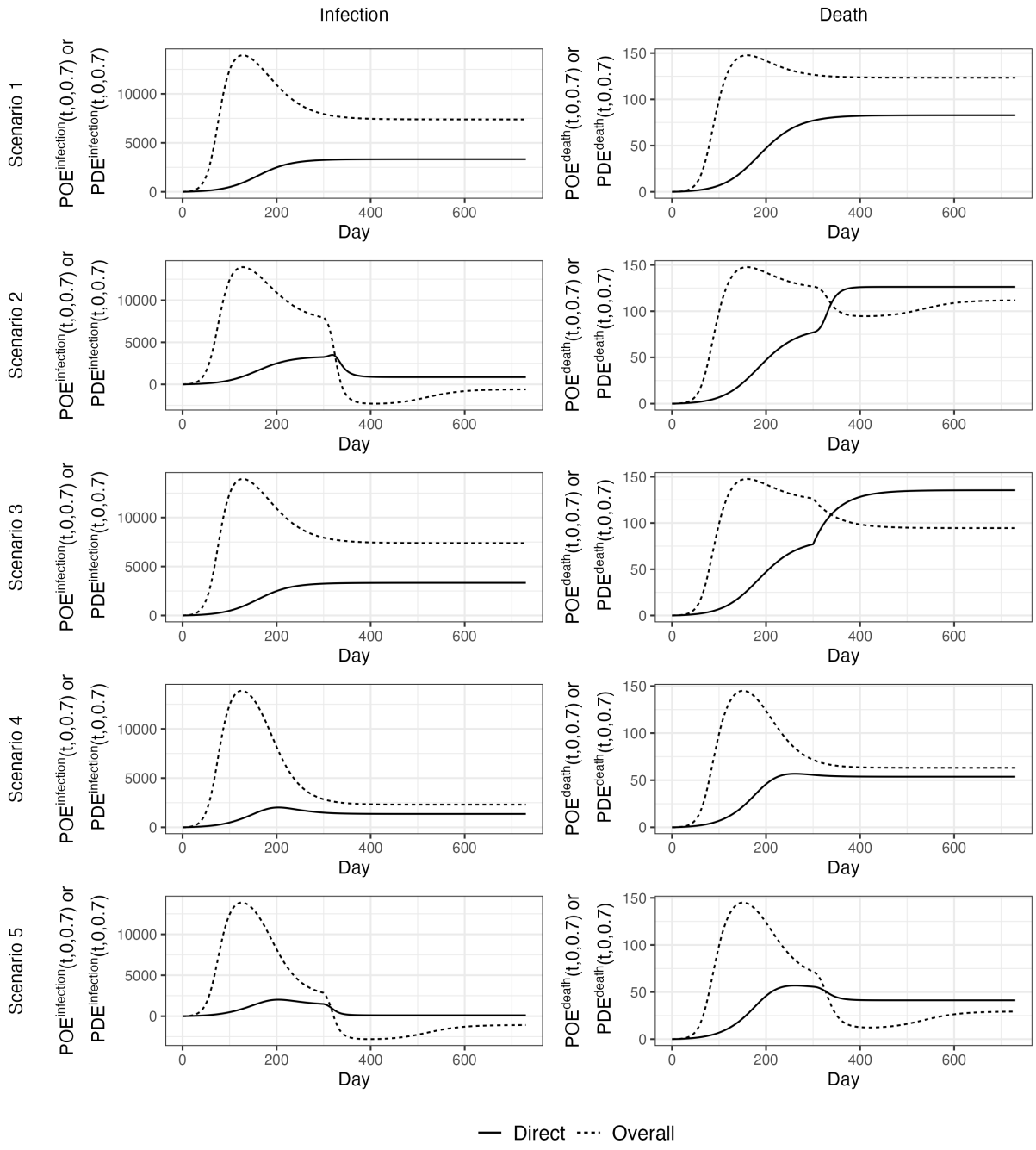

FIGURE 2. Population-level direct and overall effects given $\alpha_0 = 0$ and $\alpha_1 = 0.7$ to verify Claim 1a. Scenario 1, all parameters are time-invariant; Scenario 2, the number of effective

contacts made by a typical infectious individual per day (β) increases from 0.15 to 0.6 at Day 300; Scenario 3, probability of infection-related death (μ) increases from 0.01 to 0.1 at Day 300; Scenario 4, vaccine efficacies against infection and death start to wane linearly after Day 100 reaching 0% at Day 300; and Scenario 5, the combination of Scenarios 2 and 4.

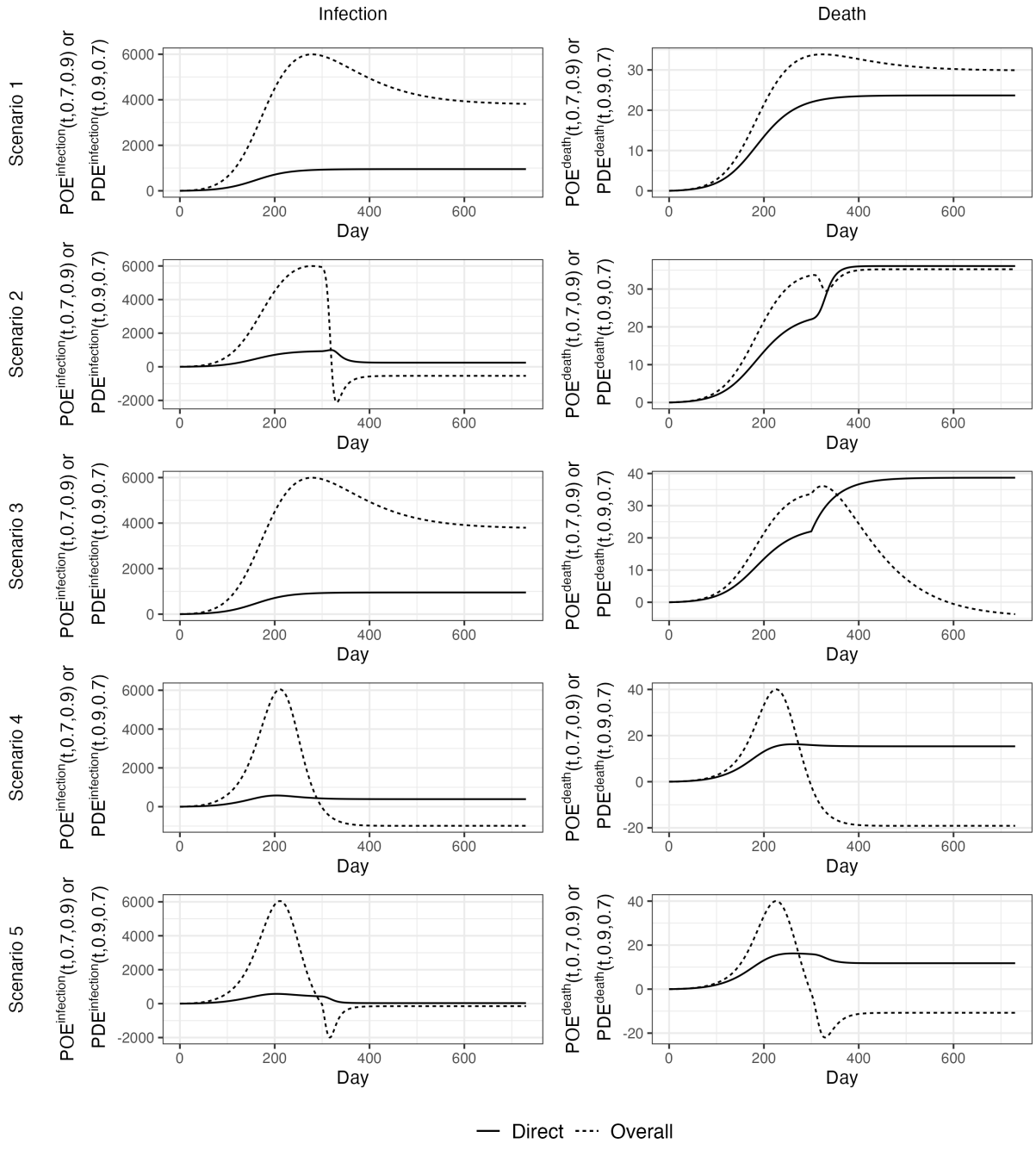

FIGURE 3. Population-level direct and overall effects given $\alpha_1 = 0.7$ and $\alpha_2 = 0.9$ to verify Claim 1b. Scenario 1, all parameters are time-invariant; Scenario 2, the number of effective contacts made by a typical infectious individual per day (β) increases from 0.15 to 0.6 at Day 300; Scenario 3, probability of infection-related death (μ) increases from 0.01 to 0.1 at Day 300; Scenario 4, vaccine efficacies against infection and death start to wane linearly after Day 100 reaching 0% at Day 300; and Scenario 5, the combination of Scenarios 2 and 4.

6. DISCUSSION

Motivated by recent research on estimating deaths averted by COVID-19 vaccination, this study defines population-level analogs of direct and overall effect estimands for infectious disease interventions, thereby enabling researchers to distinguish these estimands when conducting and interpreting related studies. Like their well-known individual-level counterparts, the population-level estimands in this study correspond to outcomes that could be observed in a two-stage randomized trial. However, the individual-level estimands correspond to differences in individual risk, whereas population-level estimands correspond to differences in counts for groups. We show the population-level direct effect can be a lower bound for population-level overall effect in a susceptible-infected-recovered-death model wherein transmission and fatality parameters do not vary and vaccine efficacies do not wane over time.

When indirect effects for the vaccinated and unvaccinated are both positive, the population-level direct effect is a lower bound of population-level overall effect (equations 10.1 and 10.2). In general, the indirect effects may be positive because vaccination reduces the infectious individuals at a given time, such that infection-naïve individuals are less likely to be infected.[20] However, as we have shown in the current study, there are conditions under which population-level direct effect

is not guaranteed to be a lower bound of overall effect due to negative indirect effects (Scenarios 2–5).

First, when the number of effective contacts made by a typical infectious individual per day (β) increases over time (Scenario 2), overall effect on infection decreases (and can be negative) because at the early stage of outbreak, the less-vaccinated group has many infected and recovered with sterilizing immunity; while the more-vaccinated group has more susceptible (i.e., infection-naïve) individuals who have escaped the earlier infections and will experience a higher force of infection at the later time (eFigure 6). β is affected by probability of transmission per contact and number of contacts per day. Transmission probability depends on meteorological factors such as absolute humidity for the influenza A virus,[21] behavioral factors such as usage of personal protective equipment,[22] and biological factors such as changes in host immunity, and evolution of strains (a proper examination of the impact of co-circulating strains on the indirect effect requires modeling the cross immunity).[23] The number of contacts can be affected by non-pharmaceutical interventions and seasonal variation in social contacts.

Second, when the probability of infection-related death (μ) increases over time (Scenario 3), overall effect on death decreases because the extensively vaccinated group(s) has more who escape the earlier infections and then experience the high fatality at the later stage of outbreak (eFigure 6). It is plausible for lethality of pathogens to increase over time: Disease severity was found to increase in the autumn wave of the 1918 flu pandemic compared to the spring-summer wave of the same pathogen in that year. Some evidence from the A/H1N1 pandemic in 2009 also suggests that lethality of a subtype may increase in the second wave compared to the first wave.[24] Increasing lethality implies that vaccination at the beginning of outbreak can lead to a negative overall effect by postponing cases. On the other hand, if the fatality rate increases with the infection peak due to

the sudden shortage of healthcare resources, the overall effect on death will be more positive because vaccines delay infection and flatten the epidemic curve. Likewise, if infection-fatality rates decline progressively due to improvements in care,[25,26] then vaccination that delays the epidemic can have amplified positive overall effects.

Third, the overall effect of increased vaccination coverage decreases and may become negative when vaccine efficacies wane (Scenario 4). In particular, the vaccination proportion of $\alpha_2 = 0.9$ may result in more infections and deaths than the proportion of $\alpha_1 = 0.7$. eFigure 7 shows the epidemic curves given $\alpha_1 = 0.7$ and $\alpha_2 = 0.9$ proportions vaccinated. When $\alpha_1 = 0.7$, the epidemic peaks earlier and slows down due to the build-up of recovered individuals with sterilizing immunity; consequently, return of full susceptibility (i.e., loss of protection) among vaccinated individuals is too late to rescue the epidemic. However, when $\alpha_2 = 0.9$, epidemic is delayed and a pool of individuals who are at risk of infection is built up, such that return of susceptibility among vaccinated individuals can rescue the epidemic.

The current study has some limitations. First, vaccination is assumed to be a one-time event at baseline before the start of outbreak, but in reality, vaccine rollout is continuous over time and may occur during outbreak. A formalization of the causal effects under a time-varying regime (e.g., by specifying a sequence of proportions vaccinated $\boldsymbol{\alpha} = (\alpha_{t_1}, \alpha_{t_2}, \dots)$ over time) will be left to future work. Second, to illustrate the counterexamples to the Claims, the current study uses a stylized model with multiple simplifications. For example, this model does not consider multiple risk groups, which may have had heterogenous susceptibility to adverse outcomes and heterogeneous mixing patterns, leading to a negative indirect effect for the unvaccinated such that the Claims do not hold. Vaccination for a subgroup could cause negative indirect effect in other subgroups by increasing risk for more severe complications. For example, empirical evidence

showed that low rubella vaccination coverage in children increased rubella incidence in the 15 years or over and incidence of congenital rubella in newborns.[27] Mathematical modelling suggested that infant vaccination against varicella could increase the rate of reactivation (i.e., zoster) in the entire population.[28,29] The examples suggest that the potential for negative indirect effects may be greater when population subgroups differ in important properties, such as (in the rubella and varicella examples) risk for adverse outcomes.

Another limitation of our model is that in the scenarios considered, changes in lethality and transmission were assumed to occur at a fixed time, whereas in reality they might well occur either in response to pathogen evolution[30] or to behavioral changes that are affected by the epidemic trajectory. However, our goal was not to describe the details of a particular epidemic but to define qualitatively the requirements for which direct effects are lower bound on overall effects at the population level. Finally, we did not consider possible adverse events after vaccination, although adverse events have important policy implications. Our focus here is on defining estimands for the protective effect of interventions on averting disease outcomes. Future studies can extend the estimands to investigate adverse events due to interventions.

In conclusion, causal estimands for effects of infectious disease interventions are originally defined at the individual level in terms of differences in individual risk under different treatment conditions, but we defined analogs of these effects at the population level. The aforementioned classification of estimands improves the conduct and interpretation of studies in evaluating the public health impact of any infectious disease intervention including vaccination. We showed—by proof for the full epidemic and by simulation for time points in the midst of the epidemic—that when transmission and fatality parameters do not vary and vaccine efficacies do not wane over time, population-level direct effect among the vaccinated (or unvaccinated) individuals can be

interpreted as a lower bound of population-level overall effect quantifying vaccine-averted (or avertible) deaths among both vaccinated and unvaccinated individuals.

REFERENCES

1. Halloran ME, Struchiner CJ. Study Designs for Dependent Happenings: *Epidemiology*. 1991;2(5):331-338. doi:10.1097/00001648-199109000-00004

2. Hudgens MG, Halloran ME. Toward Causal Inference With Interference. *Journal of the American Statistical Association*. 2008;103(482):832-842. doi:10.1198/016214508000000292

3. Vilches TN, Moghadas SM, Sah P, et al. Estimating COVID-19 Infections, Hospitalizations, and Deaths Following the US Vaccination Campaigns During the Pandemic. *JAMA Netw Open*. 2022;5(1):e2142725. doi:10.1001/jamanetworkopen.2021.42725

4. Schneider EC, Shah A, Sah P, et al. Impact of U.S. COVID-19 Vaccination Efforts: An Update on Averted Deaths, Hospitalizations, and Health Care Costs Through March 2022. To the Point. Published April 8, 2022. https://doi.org/10.26099/d3dm-fa91

5. Gavish N, Yaari R, Huppert A, Katriel G. Population-level implications of the Israeli booster campaign to curtail COVID-19 resurgence. *Sci Transl Med*. 2022;14(647):eabn9836. doi:10.1126/scitranslmed.abn9836

6. Haas EJ, McLaughlin JM, Khan F, et al. Infections, hospitalisations, and deaths averted via a nationwide vaccination campaign using the Pfizer–BioNTech BNT162b2 mRNA COVID-19 vaccine in Israel: a retrospective surveillance study. *The Lancet Infectious Diseases*. 2022;22(3):357-366. doi:10.1016/S1473-3099(21)00566-1

7. Brault A, Hart A, Uribe P, et al. Direct impact of COVID-19 vaccination in Chile: averted cases, hospitalizations, ICU admissions, and deaths. *BMC Infect Dis*. 2024;24(1):467. doi:10.1186/s12879-024-09304-1

8. Santos CVBD, Noronha TGD, Werneck GL, Struchiner CJ, Villela DAM. Estimated COVID-19 severe cases and deaths averted in the first year of the vaccination campaign in Brazil: a retrospective observational study. *The Lancet Regional Health - Americas*. 2023;17:100418. doi:10.1016/j.lana.2022.100418

9. Kayano T, Sasanami M, Kobayashi T, et al. Number of averted COVID-19 cases and deaths attributable to reduced risk in vaccinated individuals in Japan. *The Lancet Regional Health - Western Pacific*. 2022;28:100571. doi:10.1016/j.lanwpc.2022.100571

10. Jia KM, Hanage WP, Lipsitch M, et al. Estimated preventable COVID-19-associated deaths due to non-vaccination in the United States. *Eur J Epidemiol*. 2023;38(11):1125-1128. doi:10.1007/s10654-023-01006-3

11. Zhong M, Glazer T, Kshirsagar M, et al. Estimating Vaccine-Preventable COVID-19 Deaths Among Adults Under Counterfactual Vaccination Scenarios in The United States: A Modeling Study Using Observational Data. *J Pharm Pharmacol Res*. 2023;07(03). doi:10.26502/fjppr.079

12. Watson OJ, Barnsley G, Toor J, Hogan AB, Winskill P, Ghani AC. Global impact of the first year of COVID-19 vaccination: a mathematical modelling study. *The Lancet Infectious Diseases*. 2022;22(9):1293-1302. doi:10.1016/S1473-3099(22)00320-6

13. Halloran ME, Auranen K, Baird S, et al. Simulations for designing and interpreting intervention trials in infectious diseases. *BMC Med*. 2017;15(1):223. doi:10.1186/s12916-017-0985-3

14. VanderWeele TJ, Tchetgen Tchetgen EJ. Effect partitioning under interference in two-stage randomized vaccine trials. *Statistics & Probability Letters*. 2011;81(7):861-869. doi:10.1016/j.spl.2011.02.019

15. Chakladar S, Rosin S, Hudgens MG, et al. Inverse probability weighted estimators of vaccine effects accommodating partial interference and censoring. *Biometrics*. 2022;78(2):777-788. doi:10.1111/biom.13459

16. Kayano T, Nishiura H. Assessing the COVID-19 vaccination program during the Omicron variant (B.1.1.529) epidemic in early 2022, Tokyo. *BMC Infect Dis*. 2023;23(1):748. doi:10.1186/s12879-023-08748-1

17. Shoukat A, Vilches TN, Moghadas SM, et al. Lives saved and hospitalizations averted by COVID-19 vaccination in New York City: a modeling study. *The Lancet Regional Health - Americas*. 2022;5:100085. doi:10.1016/j.lana.2021.100085

18. R Core Team. R: A Language And Environment for Statistical Computing. Available at: https://www.R- project.org/ Accessed May 31, 2024.

19. FitzJohn R. odin: ODE Generation and Integration. 2024. Available at: https://github.com/mrc-ide/odin Accessed July 24, 2024.

20. Pitman R, Fisman D, Zaric GS, et al. Dynamic Transmission Modeling: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force Working Group–5. *Med Decis Making*. 2012;32(5):712-721. doi:10.1177/0272989X12454578

21. Shaman J, Kohn M. Absolute humidity modulates influenza survival, transmission, and seasonality. *Proc Natl Acad Sci USA*. 2009;106(9):3243-3248. doi:10.1073/pnas.0806852106

22. Leung NHL, Chu DKW, Shiu EYC, et al. Respiratory virus shedding in exhaled breath and efficacy of face masks. *Nat Med*. 2020;26(5):676-680. doi:10.1038/s41591-020-0843-2

23. Bushman M, Kahn R, Taylor BP, Lipsitch M, Hanage WP. Population impact of SARS-CoV-2 variants with enhanced transmissibility and/or partial immune escape. *Cell*. 2021;184(26):6229-6242.e18. doi:10.1016/j.cell.2021.11.026

24. Simonsen L, Chowell G, Andreasen V, et al. A review of the 1918 herald pandemic wave: importance for contemporary pandemic response strategies. *Annals of Epidemiology*. 2018;28(5):281-288. doi:10.1016/j.annepidem.2018.02.013

25. Horwitz LI, Jones SA, Cerfolio RJ, et al. Trends in COVID-19 Risk-Adjusted Mortality Rates. *Journal of Hospital Medicine*. 2021;16(2):90-92. doi:10.12788/jhm.3552

26. Dennis JM, McGovern AP, Vollmer SJ, Mateen BA. Improving Survival of Critical Care Patients With Coronavirus Disease 2019 in England: A National Cohort Study, March to June 2020. *Critical Care Medicine*. 2021;49(2):209-214. doi:10.1097/CCM.0000000000004747

27. Panagiotopoulos T, Antoniadou I, Valassi-Adam E. Increase in congenital rubella occurrence after immunisation in Greece: retrospective survey and systematic review. *BMJ*. 1999; 319(7223): 1462–1467.

28. Brisson M, Edmunds WJ, Gay NJ. Varicella vaccination: Impact of vaccine efficacy on the epidemiology of VZV. *J Med Virol*. 2003;70(S1):S31-S37. doi:10.1002/jmv.10317

29. Brisson M, Gay NJ, Edmunds WJ, Andrews NJ. Exposure to varicella boosts immunity to herpes-zoster: implications for mass vaccination against chickenpox. *Vaccine*. 2002;20(19-20):2500-2507. doi:10.1016/S0264-410X(02)00180-9

30. Markov PV, Ghafari M, Beer M, et al. The evolution of SARS-CoV-2. *Nat Rev Microbiol*. 2023;21(6):361-379. doi:10.1038/s41579-023-00878-2