Causal Estimands for Analyses of Averted and Avertible Outcomes due to Infectious Disease 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 M 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

During the coronavirus disease (COVID-19) pandemic, researchers attempted to estimate the number of averted and avertible outcomes due to non-pharmaceutical interventions or vaccination campaigns to quantify public health impact. However, the estimands used in these analyses have not been previously formalized. It is also unclear how these analyses relate to the broader framework of direct, indirect, total, and overall causal effects of an intervention under interference. In this study, using potential outcome notation, we adjust the direct and overall effects to accommodate analyses of averted and avertible outcomes. We use this framework to interrogate the commonly-held assumption in empirical studies that vaccine-averted outcomes via direct impact among vaccinated individuals (or vaccine-avertible outcomes via direct impact among unvaccinated individuals) is a lower bound on vaccine-averted (or -avertible) outcomes overall. To do so, we describe a susceptible-infected-recovered-death model stratified by vaccination status. When vaccine efficacies wane, the lower bound fails for vaccine-avertible outcomes. When transmission or fatality parameters increase over time, the lower bound fails for both vaccine-averted and -avertible outcomes. Only in the simplest scenario where vaccine efficacies, transmission, and fatality parameters are constant over time, outcomes averted via direct impact among vaccinated individuals (or outcomes avertible via direct impact among unvaccinated individuals) is shown to be a lower bound on overall impact on vaccine-averted (or -avertible) outcomes. In conclusion, the lower bound can fail under common violations to assumptions on constant vaccine efficacy, pathogen properties, or behavioral parameters over time. In real data analyses, estimating what seems like a lower bound on overall impact through estimating direct impact may be inadvisable without examining the directions of indirect effects. By classifying

estimands for averted and avertible outcomes and examining their relations, this study improves conduct and interpretation of research evaluating impact of infectious disease interventions.

RUNNING HEAD: Impact estimands for averted and avertible outcomes

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

1. INTRODUCTION

Recently, researchers have estimated the number of COVID-19 deaths (or infections) averted by vaccination campaigns in the United States,[1,2] Israel,[3,4] Chile,[5] Brazil,[6] and Japan.[7] Similarly, other studies, including our own,[8] have attempted to estimate vaccine-avertible deaths—defined as the number of deaths that could have been averted by vaccination, but were not because of a failure to vaccinate the unvaccinated.[8,9] Most empirical studies quantify direct protection conferred by vaccination among vaccinated individuals, but they typically assume the overall impact of vaccination is larger due to indirect protection.[4–6] Indeed, these studies often claim that their estimates represent a lower bound on overall impact of vaccination, although this claim has not been carefully verified.

To support this claim, many of these studies draw, either implicitly or explicitly, on the causal effect framework developed by Halloran and Struchiner,[10] in which the overall effect of vaccines can be decomposed into direct and indirect components.[11] However, the Halloran and Struchiner framework has not yet been formally extended to cover the specific estimands targeted in vaccine-averted and avertable analyses, which estimate *impact*. Existing effect estimands are defined by contrasts of individual risk[12,13] and have been used to estimate vaccine efficacy in clinical trials.[14,15] However, observational studies, post-licensure studies, and policy-makers are often equally or more interested in quantifying public health *impact* of vaccination in terms of counts[16] such as the number of infections, hospitalizations, and deaths in a group of individuals, instead of risk in each individual.

Motivated by recent empirical studies on vaccine-averted and -avertible COVID-19 deaths, this article seeks to fill these gaps by: (1) Clearly defining *impact* estimands as corollaries of direct and overall effect estimands for averted and avertible outcomes (i.e., counts), and (2) determining

the conditions under which direct impact is a lower bound on overall impact. To ground our discussion, we introduce a susceptible-infected-recovered-death (SIRD) model stratified by vaccination status to investigate direct and overall impact of vaccination under different scenarios.

Section 2 summarizes the definitions of direct, indirect, total, and overall effects introduced by Hudgens and Halloran,[12] and provides an alternative partitioning of overall effect into components that will align better with the estimands targeted by vaccine-averted and avertible analyses. Section 3 proposes count outcome corollaries for direct and overall effects, shows how they map onto estimands for vaccine-averted and -avertible analyses, and formalizes the claim that direct impact constitutes a lower bound on overall impact. Section 4 outlines a transmission model to simulate vaccine-averted and -avertible outcomes. Section 5 examines the conditions under which outcomes averted via direct impact among vaccinated individuals (or outcomes avertible via direct impact among unvaccinated individuals) is or is not a lower bound on overall impact on vaccine-averted (or -avertible) outcomes.

2. DIRECT, INDIRECT, OVERALL, AND TOTAL EFFECTS

Hudgens and Halloran[12] previously defined causal estimands for direct, indirect, overall, and total effects in the two-stage randomized trial, as summarized below.

Setup and notation

Consider a two-stage randomized trial with m groups indexed by $i = 1, \dots, m$, such that each group consists of N individuals indexed by $j = 1, \dots, N$ with a large group size N. All groups are assumed to be of same size N. Partial interference is assumed: Individuals make contacts 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 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.[11] 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 $Y_{ij}(\mathbf{a_i})$ denote the potential binary outcome for individual j in group i with allocation program $\mathbf{a_i}$ and let $Y_{ij}(\mathbf{a_{i,-j}}, a)$ denote the potential binary outcome when individual j has vaccination status a and the rest of group i has vaccination status $\mathbf{a_{i,-j}}$.

Individual, group, and population average potential outcomes

Hudgens and Halloran[12] define *marginal individual average potential outcome* as

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

and *individual average potential outcome*[12] as:

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

where $\Pr_{\alpha}(\cdot)$ is the probability distribution of vaccine allocation program $\boldsymbol{A_i}$ with parameter $\alpha \in [0,1]$ representing the proportion vaccinated within group i . Specifically, $\Pr_{\alpha}(\cdot)$ is the probability distribution of $\boldsymbol{A_i}$ conditional on $\sum_j^N A_{ij} = \alpha \cdot N$. Note here we use type A parameterization, which gives same effect definitions as type B parameterization suggested by VanderWeele and Tchetgen Tchetgen[11] when N is large. Definitions of $\Pr_{\alpha}(\cdot)$, type A, and type B parameterizations are given in eAppendix 1. Hudgens and Halloran[12] further define (marginal) group average potential outcomes $\bar{Y}_i(\alpha) \equiv \sum_{j=1}^N \bar{Y}_{ij}(\alpha)/N$ and $\bar{Y}_i(a; \alpha) \equiv \sum_{j=1}^N \bar{Y}_{ij}(a; \alpha)/N$ by averaging over

individuals within groups. They also define (marginal) population average potential outcomes[12] $\bar{Y}(\alpha) = \sum_{i=1}^{m} \bar{Y}_i(\alpha)\,/m$ and $\bar{Y}(a;\alpha) = \sum_{i=1}^{m} \bar{Y}_i(a;\alpha)\,/m$.

Population Average Direct, Indirect, Overall, and Total Causal Effects

Hudgens and Halloran define *population average direct casual effect*[12] as $DE(\alpha) = \bar{Y}(0;\alpha) - \bar{Y}(1;\alpha)$, comparing $\bar{Y}$ when an individual is unvaccinated versus when vaccinated, holding fixed proportion vaccinated (α). They define *population average indirect casual effect*[12] as $IE^{Unvax}(\alpha,\alpha') = \bar{Y}(0;\alpha) - \bar{Y}(0;\alpha')$, comparing $\bar{Y}$ for an unvaccinated individual in a group with α proportion vaccinated versus with α', hereafter referred to as indirect effect for the unvaccinated. As suggested previously,[17] indirect effect can be analogously defined for a vaccinated individual: $IE^{Vax}(\alpha,\alpha') = \bar{Y}(1;\alpha) - \bar{Y}(1;\alpha')$, hereafter referred to as indirect effect for the vaccinated. They also define *population average total causal effect*[12] as $TE(\alpha,\alpha') = \bar{Y}(0;\alpha) - \bar{Y}(1;\alpha')$, comparing $\bar{Y}$ an individual when they are unvaccinated in a group with α proportion vaccinated versus when they are vaccinated in a group with α'. Following from the definition, $TE(\alpha,\alpha')$ can be decomposed as $DE(\alpha') + IE^{Unvax}(\alpha,\alpha')$ or $DE(\alpha) + IE^{Vax}(\alpha,\alpha')$. Finally, they define *population average overall casual effect*[12] as $OE(\alpha,\alpha') = \bar{Y}(\alpha) - \bar{Y}(\alpha')$, comparing $\bar{Y}$ for a typical individual in a group with α proportion vaccinated versus with α'.

Overall Effect Partitioning

Previously, Hudgens and Halloran[12] showed that when $\alpha = 0$, $OE(\alpha,\alpha')$ is the weighted sum $\alpha' \cdot TE(\alpha,\alpha') + (1-\alpha') \cdot IE^{Unvax}(\alpha,\alpha')$. Here, to establish direct impact as a lower bound on overall impact (Section 3), we generalize $OE(\alpha,\alpha')$ partitioning to any $\alpha' > \alpha$ in Theorem 1.

Theorem 1 (overall effect partitioning)

$$OE(\alpha,\alpha') = \alpha \cdot IE^{Vax}(\alpha,\alpha') + (\alpha'-\alpha) \cdot TE(\alpha,\alpha') + (1-\alpha') \cdot IE^{Unvax}(\alpha,\alpha').$$

Theorem 1 is proved in eAppendix 2 and graphically illustrated in Figure 1. Theorem 1 expresses $OE(\alpha, \alpha')$ as a weighted average of three effects: 1) $IE^{Vax}(\alpha, \alpha')$, 2) $TE(\alpha, \alpha')$, and 3) $IE^{Unvax}(\alpha, \alpha')$. Intuitively, if individuals are classified by vaccination status under a pair of counterfactuals wherein the group has α or α' proportion vaccinated ($\alpha' > \alpha$), then $IE^{Vax}(\alpha, \alpha')$ is in operation for proportion α of individuals who are vaccinated under both counterfactuals; $TE(\alpha, \alpha')$ is in operation for proportion $\alpha' - \alpha$ of individuals who are vaccinated under α' but not α; and $IE^{Unvax}(\alpha, \alpha')$ is in operation for proportion $1 - \alpha'$ of individuals who are unvaccinated under both counterfactuals.

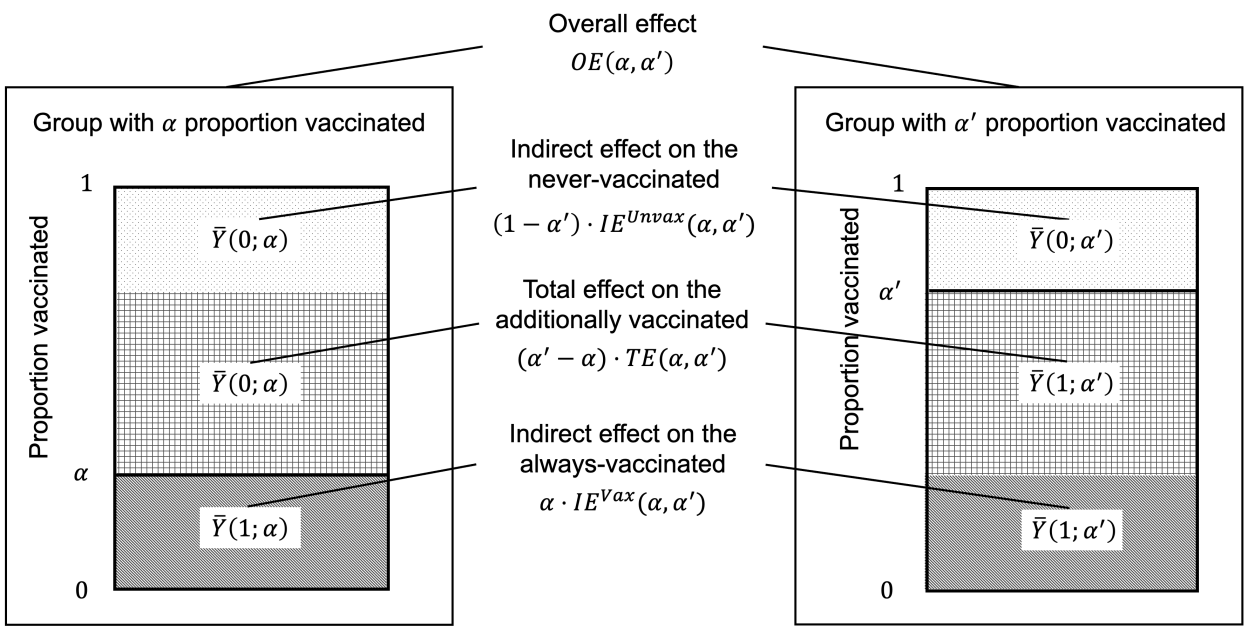


Figure 1. Graphical illustration on partitioning overall effect. The two rectangles represent a pair of counterfactuals wherein the group has α or α' proportion vaccinated ($\alpha' > \alpha$). Individuals fall into three categories based on their vaccination status under the counterfactuals: 1) The dotted region represents those who are unvaccinated under both counterfactuals (hereafter referred to as "never-vaccinated") and for whom $IE^{Unvax}(\alpha, \alpha')$ is in operation; 2) the gridded region represents those who are unvaccinated under α but vaccinated under α' and for whom $TE(\alpha, \alpha')$ 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 $IE^{Vax}(\alpha, \alpha')$ is in operation. Theorem 1 shows that $OE(\alpha, \alpha')$ is a weighted average of three effects: 1) $IE^{Unvax}(\alpha, \alpha')$, 2) $TE(\alpha, \alpha')$, and 3) $IE^{Vax}(\alpha, \alpha')$, each weighted by the proportion of individuals for whom the effect is in operation respectively: 1) $1 - \alpha'$ for the never-vaccinated, 2) $\alpha' - \alpha$ for the additionally-vaccinated, and 3) α for the always-vaccinated.

3. ESTIMANDS FOR ANALYSES OF AVERTED AND AVERTIBLE OUTCOMES

Up until now, all effect estimands (except indirect effect for vaccinated)[17] have been previously defined by Hudgens and Halloran in a two-stage randomized trial. To define estimands for analyses of averted and avertible outcomes, we now expand on the terminology of two-stage randomization. Although a two-stage randomized trial is rarely conducted in practice, our goal is to be explicit about the definition of estimands in recent observational studies on averted and avertible outcomes, and to place these observational studies in the context of target trial emulation, which is playing a growing role in causal inference.

The original overall effect (OE) and direct effect (DE) are defined for individual risk,[10] and their magnitudes are not comparable because OE and DE use different denominators. However, empirical observational studies of averted and avertible deaths use count outcomes, assuming vaccination prevents more deaths overall than directly (i.e., OE multiplied by total population size is greater than multiplying DE with number of vaccinated individuals in the presence of indirect protection for unvaccinated individuals).[4–9] To verify this assumption, Section 3 defines two estimands for averted and avertible outcomes—direct impact and overall impact—that align with estimands targeted by literature on vaccine-averted and avertible outcomes.[1–9] Sections 4 and 5 then verify conditions under which direct impact is or is not a lower bound on overall impact.

This paper adds to existing literature by: (1) Introducing an additional time index t to examine outcomes at multiple timepoints post-vaccination, (2) defining direct and overall impact (i.e., corollaries of direct and overall effect estimands for count outcomes) to be used in averted and avertible outcome analyses, (3) defining the relationships between direct and overall impact, and (4) most importantly, using analytical and simulation approaches to identify conditions under which direct impact may or may not be a lower bound on overall impact.

Notation

To examine direct and overall impact over time, we incorporate time index t so that $\bar{Y}(t, a; \alpha)$ denotes population average cumulative incidence of having developed the outcome by time t, and similarly for $\bar{Y}(t; \alpha)$ and all the effect estimands in Section 2. $t = 0$ denotes start of follow-up. In principle, t can be of any timescale, while t is in days in the simulations in Section 5.

Now let α_1 denote a particular vaccination proportion chosen to be implemented in the group, let α_2 denote some hypothetical higher proportion (i.e., $\alpha_2 > \alpha_1$), and let α_0 denote some hypothetical lower proportion (i.e., $\alpha_0 < \alpha_1$).

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.[1–7,18–20] Vaccine-averted infections (or deaths) is an impact estimand based on the causal question: How many infections (or deaths) have been averted under the particular proportion vaccinated (α_1) compared to the counterfactual in the absence of vaccination ($\alpha_0 = 0$)?

Alternatively, researchers have also estimated the vaccine-avertible deaths—those that could have been averted by vaccination but were not because of a failure to vaccinate.[8,9] The causal

question is: How many infections (or deaths) could have been averted under full vaccination ($\alpha_2 = 1$), but were not averted given the particular proportion vaccinated (α_1)?

In general, estimands are defined by comparing number of infections (or deaths) in a typical group under the particular proportion vaccinated (α_1) versus a lower (α_0) or higher (α_2) proportion.[9] We term these as *impact* estimands because previous literature has referred to vaccine-averted infections (or deaths) as (population) impact of vaccination.[5,7,18,19]

Overall Impact

For $\alpha_0 < \alpha_1$, overall impact is defined as:

$$\delta^O(t, \alpha_0, \alpha_1) = N \cdot OE(t, \alpha_0, \alpha_1). \quad (1.1)$$

For $\alpha_2 > \alpha_1$, overall impact is defined as:

$$\delta^O(t, \alpha_1, \alpha_2) = N \cdot OE(t, \alpha_1, \alpha_2). \quad (1.2)$$

Overall impact directly answers the two aforementioned causal questions on quantifying vaccine-averted and -avertible outcomes. For reasons clarified later, we force overall impact to have the same sign when α_1 is compared to a lower (α_0) or higher (α_2) hypothetical value.

Note mathematical modelling studies implicitly refer to overall impact when estimating vaccine-averted deaths by simulating the epidemic trajectory under a hypothetical proportion vaccinated (e.g., $\alpha_0 = 0$) and comparing it with the trajectory under the particular vaccination campaign (α_1).

Furthermore, by Theorem 1 and $TE(t, \alpha_0, \alpha_1) = DE(t, \alpha_1) + IE^{Unvax}(t, \alpha_0, \alpha_1)$, we decompose:

$$\delta^O(t,\alpha_0,\alpha_1) = N \cdot OE(t,\alpha_0,\alpha_1)$$
$$= N \cdot \alpha_0 \cdot IE^{Vax}(t,\alpha_0,\alpha_1) + N \cdot (\alpha_1 - \alpha_0) \cdot \left(DE(t,\alpha_1) + IE^{Unvax}(t,\alpha_0,\alpha_1)\right)$$
$$+ N \cdot (1-\alpha_1) \cdot IE^{Unvax}(t,\alpha_0,\alpha_1)$$
$$= N \cdot \alpha_0 \cdot IE^{Vax}(t,\alpha_0,\alpha_1) + N \cdot (\alpha_1 - \alpha_0) \cdot DE(t,\alpha_1) + N \cdot (1-\alpha_0)$$
$$\cdot IE^{Unvax}(t,\alpha_0,\alpha_1) \quad (2.1)$$

for $\alpha_1 > \alpha_0$. The first term on the right-hand side of last line of equation (2.1) scales $IE^{Vax}(t,\alpha_0,\alpha_1)$ by number vaccinated under α_0. The second term scales $DE(t,\alpha_1)$ by the additional number vaccinated under α_1 compared to α_0. The third term scales $IE^{Unvax}(t,\alpha_0,\alpha_1)$ by number unvaccinated under α_0.

Similarly, by Theorem 1 and $TE(t,\alpha_1,\alpha_2) = DE(t,\alpha_1) + IE^{Vax}(t,\alpha_1,\alpha_2)$, we decompose:

$$\delta^O(t,\alpha_1,\alpha_2) = N \cdot OE(t,\alpha_1,\alpha_2)$$
$$= N \cdot \alpha_2 \cdot IE^{Vax}(t,\alpha_1,\alpha_2) + N \cdot (\alpha_2 - \alpha_1) \cdot DE(t,\alpha_1) + N \cdot (1-\alpha_2)$$
$$\cdot IE^{Unvax}(t,\alpha_1,\alpha_2) \quad (2.2)$$

for $\alpha_2 > \alpha_1$. The first term on the right-hand side of last line of equation (2.2) scales $IE^{Vax}(t,\alpha_0,\alpha_1)$ by number vaccinated under α_2. The second term scales $DE(t,\alpha_1)$ by the additional number vaccinated under α_2 compared to α_1. The third term scales $IE^{Unvax}(t,\alpha_0,\alpha_1)$ by number unvaccinated under α_2.

Direct Impact

Because overall impact (δ^O) contrasts two different vaccination proportions, at minimum researchers need to observe two noninteracting groups where each is "assigned" to one of the vaccination strategies to estimate overall impact.[12] Ideally, vaccination would be assigned via a two-stage randomized trial. However, these trials are rare because they are expensive and hard to justify.[21] Instead, most empirical studies only observe a single group under one vaccination

proportion (α_1), in which case researchers can only estimate an impact corollary of direct effect. Examples include observational studies comparing discrete hazards or incidence rates between vaccinated and unvaccinated individuals based on national vaccine data systems.[4–9] These studies multiply direct effect with number vaccinated (or unvaccinated) to estimate outcomes averted (or avertible) via direct impact among vaccinated (or unvaccinated) individuals,[4–9] and then generally assume that it is lower than overall impact in the entire population.[4–6,8] Here, we formalize this definition of direct impact and show how it relates to overall impact.

Some empirical observational studies have estimated deaths averted via direct impact among vaccinated individuals[4–7] using formulas similar to $N \cdot (\alpha_1 - \alpha_0) \cdot DE(t, \alpha_1)$ by setting $\alpha_0 = 0$, while others have estimated deaths avertible via direct impact among unvaccinated individuals[8,9] using formulas similar to $N \cdot (\alpha_2 - \alpha_1) \cdot DE(t, \alpha_1)$ by setting $\alpha_2 > \alpha_1$ (Note these studies have also considered increases in proportion vaccinated over time, such as under a vaccine rollout, which here we ignore for simplicity). Following the literature,[4–9,18] let direct impact (δ^D) for any $\alpha, \alpha' \in [0,1]$ be $\delta^D(t, \alpha, \alpha') = N \cdot |\alpha' - \alpha| \cdot DE(t, \alpha')$.

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

$$\delta^D(t, \alpha_0, \alpha_1) = N \cdot (\alpha_1 - \alpha_0) \cdot DE(t, \alpha_1) \quad (3.1)$$

and for $\alpha_2 > \alpha_1$,

$$\delta^D(t, \alpha_2, \alpha_1) = N \cdot (\alpha_2 - \alpha_1) \cdot DE(t, \alpha_1). \quad (3.2)$$

Note $\delta^D(t, \alpha_0, \alpha_1)$ or $\delta^D(t, \alpha_2, \alpha_1)$ is not a meaningful causal estimand by itself because $DE(t, \alpha_1)$ is conditional on α_1 only and does not account for change in DE when proportion vaccinated is α_0 or α_2 instead of α_1. Importantly, now direct impact can be a lower bound on overall impact. This is because Theorem 1 decomposes OE into TE, IE^{Unvax}, and IE^{Vax}, such

that we can map direct impact onto $N \cdot |\alpha' - \alpha| \cdot DE(t, \alpha')$ in equations 2.1 and 2.2, which are now written as:

$$\delta^O(t, \alpha_0, \alpha_1) = N \cdot \alpha_0 \cdot IE^{Vax}(t, \alpha_0, \alpha_1) + \delta^D(t, \alpha_0, \alpha_1) + N \cdot (1 - \alpha_0) \cdot IE^{Unvax}(t, \alpha_0, \alpha_1) \quad (4.1)$$

for $\alpha_1 > \alpha_0$, and

$$\delta^O(t, \alpha_1, \alpha_2) = N \cdot \alpha_2 \cdot IE^{Vax}(t, \alpha_1, \alpha_2) + \delta^D(t, \alpha_2, \alpha_1) + N \cdot (1 - \alpha_2) \cdot IE^{Unvax}(t, \alpha_1, \alpha_2) \quad (4.2)$$

for $\alpha_2 > \alpha_1$. Since overall impact has the same sign when α_1 is compared to a lower (α_0) or higher (α_2) hypothetical value, we formalize the assumption often made in the empirical vaccine-averted and -avertible outcome studies—that is, direct impact is a lower bound on overall impact by considering Claim 1:

$$\delta^O(t, \alpha_0, \alpha_1) \geq \delta^D(t, \alpha_0, \alpha_1) \qquad \forall\, \alpha_0 < \alpha_1,$$

and

$$\delta^O(t, \alpha_1, \alpha_2) \geq \delta^D(t, \alpha_2, \alpha_1) \qquad \forall\, \alpha_2 > \alpha_1.$$

If Claim 1 is true, direct impact, which can be estimated using commonly available data,[4–9,18] is a lower bound on overall impact that is relevant for policy-making and retrospective policy evaluation requiring samples from a population of groups as in two-stage randomized trials.[12]

Following the literature, we consider two special cases of Claim 1. In Claim 1a, a particular vaccination proportion α_1 is compared to a hypothetical of no vaccination ($\alpha_0 = 0$) to quantify vaccine-averted outcomes.

Claim 1a (vaccine-averted outcomes):

$$\delta^O(t, 0, \alpha_1) \geq \delta^D(t, 0, \alpha_1).$$

In words, Claim 1a asserts that for $\alpha_1 > \alpha_0 = 0$, outcomes averted via direct impact of current vaccination among vaccinated individuals is a lower bound on total vaccine-averted outcomes among both vaccinated and unvaccinated individuals.

In Claim 1b, α_1 is compared to a hypothetical of near-universal vaccination ($\alpha_2 = 0.9 > \alpha_1$) to quantify vaccine-avertible outcomes, following the literature.[9]

Claim 1b (vaccine-avertible outcomes):

$$\delta^O(t, \alpha_1, 0.9) \geq \delta^D(t, 0.9, \alpha_1).$$

In words, Claim 1b asserts that for $\alpha_2 = 0.9 > \alpha_1$, outcomes avertible via direct impact of current vaccination among some unvaccinated individuals is a lower bound on vaccine-avertible outcomes among both vaccinated and unvaccinated individuals. Note we do not compare to $\alpha_2 = 1$ because DE would be undefined under full vaccination, and there would not be an epidemic under full vaccination with a highly effective vaccine.

Based on the partitioning in equations 4.1 and 4.2, Claims 1a and 1b hold if IE^{Vax} and IE^{Unvax} are non-negative. However, it is not immediately intuitive when that occurs. Therefore, we describe a transmission model to check Claims 1a and 1b under various scenarios (Section 4), and then describe conditions under which direct impact is or is not a lower bound on overall impact (Section 5).

4. TRANSMISSION MODEL

The SIRD Model with Vaccination at Baseline

An SIRD model represents a well-mixed group in a two-stage randomized trial assuming partial interference.[12] To simulate direct and overall impact, we simulate a typical group with a large size under a pair of counterfactual vaccination proportions. In this simulation, (marginal)

group average potential outcome is equivalent to (marginal) population average potential outcome because there is only one group in the population. 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 is randomly assigned to a vaccination policy wherein the proportion to be vaccinated is α, and individuals are randomly assigned a vaccination status a at baseline ($a = 1$ for vaccinated and $a = 0$ for unvaccinated; for equation [5], subscript v denotes vaccinated and u for unvaccinated) based on their group proportion. The vaccine is "leaky" in protecting against infection and infection-related death—that is, vaccination reduces susceptibility by θ against infection (i.e., vaccine efficacy against infection [$VE_{infection}$] is $(1-\theta)$ • 100%) and reduces susceptibility by κ against death (i.e., vaccine efficacy against death given infection [$VE_{death|infection}$] is $(1-\kappa)$ • 100%). Individuals mix homogeneously such that each vaccinated or unvaccinated susceptible individual is equally likely to contact any infectious individual. Vaccinated and unvaccinated infectious individuals are equally contagious. The transmission dynamics are:

$$\left.\begin{array}{c} dS_{u,\alpha}(t)/dt = -\lambda_\alpha(t)\cdot S_{u,\alpha}(t) \\ dS_{v,\alpha}(t)/dt = -\theta\cdot\lambda_\alpha(t)\cdot S_{v,\alpha}(t) \\ dI_{u,\alpha}(t)/dt = \lambda_\alpha(t)\cdot S_{u,\alpha}(t) - \gamma\cdot I_{u,\alpha}(t) \\ dI_{v,\alpha}(t)/dt = \theta\cdot\lambda_\alpha(t)\cdot S_{v,\alpha}(t) - \gamma\cdot I_{v,\alpha}(t) \\ dR_{u,\alpha}(t)/dt = (1-\mu)\cdot\gamma\cdot I_{u,\alpha}(t) \\ dR_{v,\alpha}(t)/dt = (1-\kappa\cdot\mu)\cdot\gamma\cdot I_{v,\alpha}(t) \\ dD_{u,\alpha}(t)/dt = \mu\cdot\gamma\cdot I_{u,\alpha}(t) \\ dD_{v,\alpha}(t)/dt = \kappa\cdot\mu\cdot\gamma\cdot I_{v,\alpha}(t) \end{array}\right\} \quad (5)$$

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 (5), $S_{u,\alpha}(t)$ 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 parameter values used in simulation.

Software

All simulations and visualization are conducted using R 4.2.2 (R Foundation for Statistical Computing, Vienna, Austria).[22] All models are implemented using R package *odin*.[23] Code is available at https://github.com/katjia/impact_estimands.

5. WHEN IS DIRECT IMPACT A LOWER BOUND ON OVERALL IMPACT?

Rationale

The SIRD model has many simplifying assumptions compared to real-world settings. However, if one can show a counterexample to Claim 1 based on the simplest SIRD model, then Claim 1 is not guaranteed to be true in more general and realistic models. To identify counterexamples of Claim 1, we consider common violations to assumptions on time-invariant parameters so that indirect effects can be negative: (1) Number of effective contacts may increase over time due to meteorological factors,[24] lifting of non-pharmaceutical interventions,[25] and seasonal variation in social contacts, (2) infection-fatality risk may increase over time,[26] (3) waning immunity clearly occurred in the Delta and Omicron waves of COVID-19 pandemic.[27,28]

Scenarios

We check the Claims under 5 scenarios (Table). Scenario 1 refers to the SIRD model in equation (5) with time-invariant parameters. eTable 1 lists model parameters for which one or two parameters vary under each Scenario separately: Scenario 2 increases 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), in eAppendix 4, we prove that Claim 1 (i.e., direct impact is a lower bound on overall impact 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 whether 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 not under any other Scenarios. The Table summarizes the results. Figures 2 and 3 show trajectories of direct and overall impact 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, Latin hypercube sampling verifies that Claim 1 holds under Scenario 1 (eAppendix 5). Under Scenario 2 where β increases, Claims 1a and 1b do not hold: Direct impacts are not lower bounds on overall impacts (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 outcomes) holds (Figure 2). Latin hypercube sampling verifies that Claim 1a holds under Scenario 4 (eAppendix 5). However, Claim 1b (vaccine-avertible outcomes) 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.

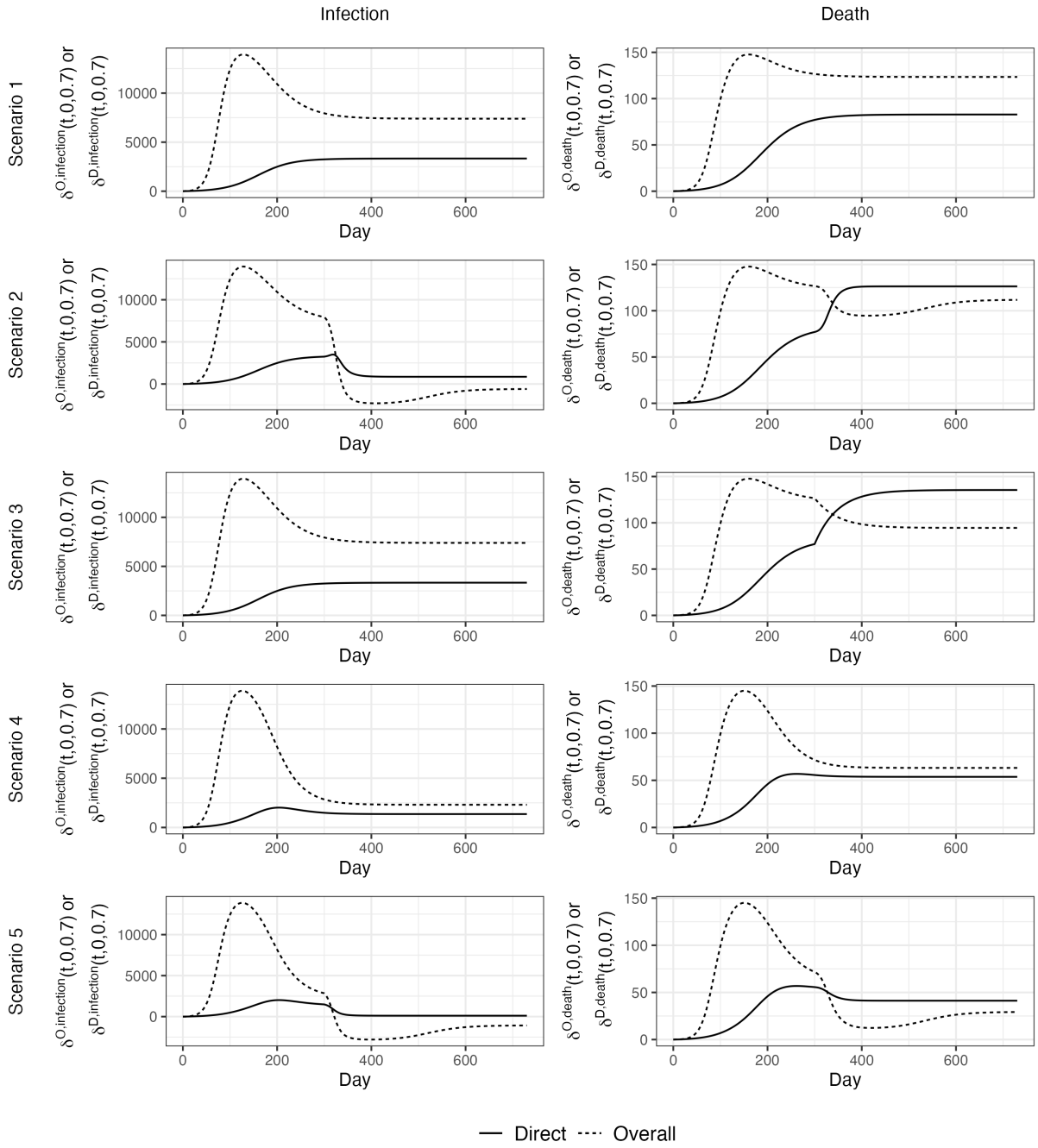


FIGURE 2. Direct impact and overall impact given $\alpha_0 = 0$ and $\alpha_1 = 0.7$ to verify Claim 1a. Scenario 1, all parameters are time-invariant; Scenario 2, the number of effective contacts made by a typical infectious individual per day (β) increases from 0.15 to 0.6 at Day 300; Scenario 3, probability 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.

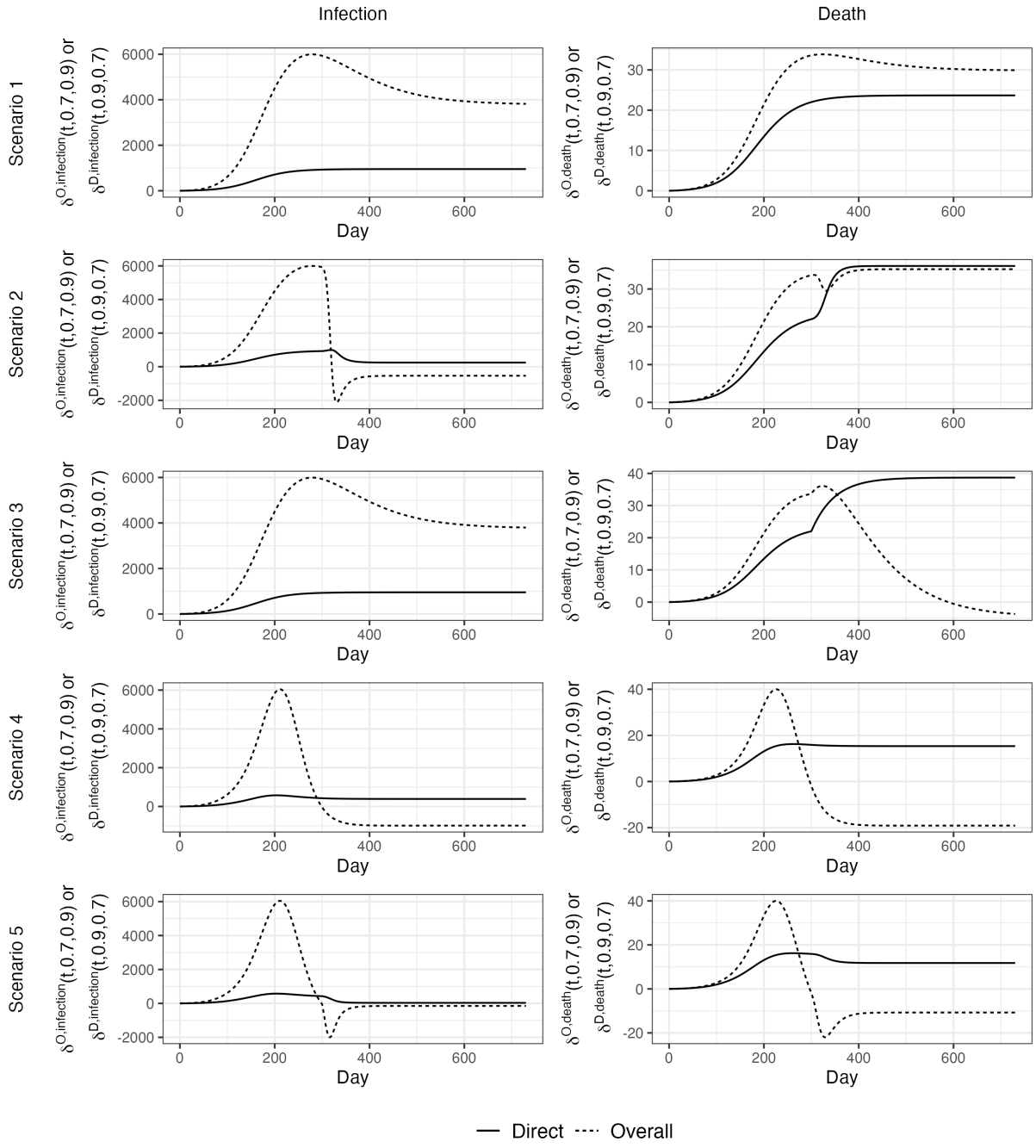


FIGURE 3. Direct impact and overall impact given $\alpha_1 = 0.7$ and $\alpha_2 = 0.9$ to verify Claim 1b. Scenario 1, all parameters are time-invariant; Scenario 2, the number of effective contacts

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 deaths averted by COVID-19 vaccination, this study adjusts estimands defined by Hudgens and Halloran to accommodate analyses on averted and avertible outcomes due to infectious disease interventions, thereby enabling researchers to distinguish the estimands when conducting and interpreting related studies. An epidemic model is simulated to verify the commonly held claim that outcomes averted (or avertible) via direct impact among the vaccinated (or unvaccinated) individuals is a lower bound on overall impact. Based on the SIRD model, the lower bound fails when transmission or fatality parameter increases, or vaccine efficacies wane, implying that the lower bound is not guaranteed to hold for more general and realistic models. Consequently, it would be overly optimistic for empirical studies[4–9] to assume that they have estimated a lower bound on true number of averted (or avertible) outcomes through estimating the direct impact without examining the directions of indirect effects.

When indirect effects for the vaccinated and unvaccinated are both positive (that is, higher vaccination coverage yields positive number of averted or avertible outcomes), direct impact is a lower bound on overall impact (equations 4.1 and 4.2). Subsequently, we show in eAppendix 4 that, at the end of epidemic, deaths averted via direct impact would be a lower bound on overall impact in a SIRD model given time-invariant parameters and vaccination at baseline, thanks to the positive indirect effects, a finding proven more recently in a preprint by Lin et al. (2024).[29] In general, 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.[30] However, as shown above, there are scenarios under which direct impact is not a lower bound on overall impact 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 impact 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 earlier infections and will experience a higher force of infection at later time (eFigure 6). β is affected by meteorological factors,[24] behavioral factors (e.g., usage of personal protective equipment),[31] biological factors (e.g., changes in host immunity, evolution of strains),[32] non-pharmaceutical interventions, and seasonal variation in social contacts.

Second, when probability of infection-related death (μ) increases over time (Scenario 3), overall impact on death decreases because the extensively vaccinated group(s) has more who escape earlier infections and then experience higher fatality at later stage of outbreak (eFigure 6). It is plausible for lethality of pathogens to increase over time: Disease severity increased in the autumn wave of the 1918 flu pandemic compared to the spring-summer wave of the same pathogen in that year.[26] Increasing lethality implies that vaccination at beginning of outbreak can decrease overall impact by postponing cases. On the other hand, if fatality rate increases with infection peak due to the sudden shortage of healthcare resources, overall impact 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,[33,34] then vaccination that delays the epidemic can have amplified positive overall impact.

Third, overall impact may become negative when vaccine efficacies wane (Scenario 4). In particular, vaccination proportion of $\alpha_2 = 0.9$ may result in more infections and deaths than the proportion of $\alpha_1 = 0.7$. eAppendix 8 discusses this scenario in greater detail.

In addition, if there were multiple risk groups with heterogenous susceptibility to adverse outcomes and heterogeneous mixing patterns, 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-and-over and incidence of congenital rubella in newborns.[35]

The current study has some limitations. First, vaccination is assumed to be a one-time event at baseline before start of outbreak, but in reality, vaccine rollout is continuous over time and may occur during outbreak. Second, throughout, we have considered the case where vaccination occurs at random. However, in most empirical settings there may be strong confounding due to staged rollout of vaccines and differences in vaccine acceptance by behavioral and health characteristics. Such confounding if uncontrolled threatens the validity of inferences about the effects, whether or not a bound is valid in ideal (unconfounded) circumstances.

Another limitation 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[36] 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 describe qualitatively conditions under which infectious disease interventions may not prevent more cases overall than directly. Finally, the SIRD model does not consider deaths due to other factors, meaning that simulations are applicable to studies whose outcome of interest is a consequence of infection. Over a short time frame (e.g., 1 year), it is acceptable to consider deaths due to infection

only, when other causes of deaths are negligible. We also did not consider possible adverse events after vaccination, although adverse events have important policy implications. Our focus here is impact of interventions on averting disease outcomes. Future studies can extend the estimands to investigate adverse events.

In conclusion, this study examines the commonly held assumption in empirical vaccine-averted and avertible analyses that direct impact is a lower bound on overall impact due to indirect protection. We show that the use of direct impact as a lower bound on overall impact is reliable only under very strong and often unrealistic assumptions. Therefore, it is inadvisable for empirical studies to assume this relation without examining the directions of indirect effects. Alternatively, if researchers want to estimate the averted and avertible outcomes, a transmission model should be used to capture the overall impact, thereby improving clarity about what is estimated despite the expense of making additional modelling assumptions.

REFERENCES

1. 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

2. 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. April 8, 2022. Accessed December 10, 2023. https://doi.org/10.26099/d3dm-fa91

3. 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

4. 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

5. 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

6. 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

7. 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

8. 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*. Published online April 24, 2023. doi:10.1007/s10654-023-01006-3

9. 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

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

11. 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

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

13. Halloran ME, Struchiner CJ, Longini IM. Study Designs for Evaluating Different Efficacy and Effectiveness Aspects of Vaccines. *American Journal of Epidemiology*. 1997;146(10):789-803. doi:10.1093/oxfordjournals.aje.a009196

14. Polack FP, Thomas SJ, Kitchin N, et al. Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine. *N Engl J Med*. 2020;383(27):2603-2615. doi:10.1056/NEJMoa2034577

15. Frenck RW, Klein NP, Kitchin N, et al. Safety, Immunogenicity, and Efficacy of the BNT162b2 Covid-19 Vaccine in Adolescents. *N Engl J Med*. 2021;385(3):239-250. doi:10.1056/NEJMoa2107456

16. Standaert B, Dort T, Toumi M. Vaccine Efficacy, Effectiveness, Or Impact: Which One To Choose In Economic Evaluations Of Vaccines? *Value in Health*. 2017;20(9):A754. doi:10.1016/j.jval.2017.08.2115

17. 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

18. 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

19. 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

20. 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

21. Magosi LE, Zhang Y, Golubchik T, et al. Deep-sequence phylogenetics to quantify patterns of HIV transmission in the context of a universal testing and treatment trial – BCPP/Ya Tsie trial. *eLife*. 2022;11:e72657. doi:10.7554/eLife.72657

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

23. FitzJohn R. odin: ODE Generation and Integration. Published online 2024. https://github.com/mrc-ide/odin

24. 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

25. Li Y, Campbell H, Kulkarni D, et al. The temporal association of introducing and lifting non-pharmaceutical interventions with the time-varying reproduction number (R) of SARS-CoV-2: a modelling study across 131 countries. *The Lancet Infectious Diseases*. 2021;21(2):193-202. doi:10.1016/S1473-3099(20)30785-4

26. 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

27. Chemaitelly H, Tang P, Hasan MR, et al. Waning of BNT162b2 Vaccine Protection against SARS-CoV-2 Infection in Qatar. *N Engl J Med*. 2021;385(24). doi:10.1056/NEJMoa2114114

28. Menegale F, Manica M, Zardini A, et al. Evaluation of Waning of SARS-CoV-2 Vaccine–Induced Immunity: A Systematic Review and Meta-analysis. *JAMA Netw Open*. 2023;6(5):e2310650. doi:10.1001/jamanetworkopen.2023.10650

29. Lin L, Demirhan H, Stone L. Evaluating the Effectiveness of Vaccination Campaigns: A Sird-Type Model Versus a Statistical Approach. Published online 2024. doi:http://dx.doi.org/10.2139/ssrn.4940607

30. 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

31. 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

32. 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

33. 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

34. 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

35. Panagiotopoulos T, Antoniadou I, Valassi-Adam E. Increase in congenital rubella occurrence after immunisation in Greece: retrospective survey and systematic review. 1999;319.

36. 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

eAppendix of "Causal Estimands for Analyses of Averted and Avertible Outcomes due to Infectious Disease Interventions"

Table of Contents

eAppendix 1. Type A versus type B parameterizations 2

eAppendix 2. Proof of Theorem 1 (overall effect partitioning) 4

eAppendix 3. The susceptible-infected-recovered-death (SIRD) model 5

eAppendix 4. Proving Claim 1 under Scenario 1 in the SIRD model at $t \to \infty$ 8

eAppendix 5. Latin hypercube sampling to verify Claim 1 under Scenario 1 and Claim 1a under Scenario 4 16

eAppendix 6. Trajectories of indirect effects for the vaccinated and unvaccinated. 19

eAppendix 7. Trajectories of proportion susceptible by vaccination status over time 22

eAppendix 8. Epidemic curves under Scenario 4. 24

References for eAppendix 25

eAppendix 1. Type A versus type B parameterizations

In the main text, the potential outcomes $\bar{Y}_{ij}(\alpha)$ and $\bar{Y}_{ij}(a;\alpha)$ are defined under type A parameterization. eAppendix 1 will define type A and B parameterizations. We assume a large group size N such that results are equivalent under both types of parameterizations.[1]

Let $\rho(\boldsymbol{A_i} = \mathbf{a_i}; \alpha)$ denote the probability that group i receive allocation program $\mathbf{a_i}$ given parameter α.

Definition 1 (A) According to VanderWeele and Tchetgen Tchetgen,[1] a parameterization is said to be of type A with parameter $\alpha = (N, K)$ for group i if the allocation program $\boldsymbol{A_i}$ is randomly allocated conditional on $\sum_{j=1}^{N} A_{ij} = K$ with probability mass function:

$$\rho(\boldsymbol{A_i}; \alpha) = I(\sum_{j=1}^{N} A_{ij} = K) / \binom{N}{K}$$

(B) A parameterization is said to be of type B with parameter α if the allocation program $\boldsymbol{A_i}$ is randomly assigned to individuals in group i according to the known Bernoulli probability mass function:

$$\rho(\boldsymbol{A_i}; \alpha) = \prod_{j=1}^{N} \alpha^{A_{ij}} (1-\alpha)^{1-A_{ij}}$$

According to VanderWeele and Tchetgen Tchetgen,[1] Hudgens and Halloran [2] defined causal effects under Type A parameterization.

Definition 2 The *marginal individual average potential outcome* of Hudgens and Halloran [2] is:

$$\bar{Y}_{ij}^{HH}(\alpha) \equiv \sum_{\mathbf{s} \in \mathcal{A}(N)} Y_{ij}(\mathbf{a}_i = \mathbf{s}) \Pr_{\alpha}(\boldsymbol{A_i} = \mathbf{s})$$

where $\Pr_{\alpha}(\boldsymbol{A_i} = \mathbf{s}) = \rho(\boldsymbol{A_i} = \boldsymbol{s}; \alpha)$ and $\rho(\boldsymbol{A_i} = \boldsymbol{s}; \alpha)$ follows Definition 1 (A) under type A parameterization. The *individual average potential outcome* [2] is:

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

where $\Pr_{\alpha}\left(\boldsymbol{A}_{\boldsymbol{i},-\boldsymbol{j}} = \boldsymbol{s} | A_{ij} = a\right) = \frac{\rho\left(\boldsymbol{A}_{\boldsymbol{i},-\boldsymbol{j}}=\boldsymbol{s}, A_{ij}=a;\alpha\right)}{\sum_{\boldsymbol{s}'\in\mathcal{A}(N-1)} \rho\left(\boldsymbol{A}_{\boldsymbol{i},-\boldsymbol{j}}=\boldsymbol{s}', A_{ij}=a;\alpha\right)}$ and $\rho\left(\boldsymbol{A}_{\boldsymbol{i},-\boldsymbol{j}} = \boldsymbol{s}, A_{ij} = a;\alpha\right)$ follows Definition 1 (A) under type A parameterization.

Alternatively, VanderWeele and Tchetgen Tchetgen[1] proposed the definitions of potential outcomes under Type B parameterization.

Definition 3 Under type B parameterization, the *marginal individual average potential outcome*[1] is:

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

where $\Pr_{\alpha}(\boldsymbol{A}_{\boldsymbol{i}} = \mathbf{s}) = \rho(\boldsymbol{A}_{\boldsymbol{i}} = \boldsymbol{s};\alpha)$ and $\rho(\boldsymbol{A}_{\boldsymbol{i}} = \boldsymbol{s};\alpha)$ follows Definition 1 (B) under type B parameterization. The *individual average potential outcome*[1] is:

$$\bar{Y}_{ij}^{B}(a;\alpha) = \sum_{\boldsymbol{s}\in\mathcal{A}(N-1)} Y_{ij}\left(\mathbf{a}_{\boldsymbol{i},-\boldsymbol{j}} = \boldsymbol{s}, a_{ij} = a\right) \Pr_{\alpha}\left(\boldsymbol{A}_{\boldsymbol{i},-\boldsymbol{j}} = \boldsymbol{s}\right)$$

where $\Pr_{\alpha}\left(\boldsymbol{A}_{\boldsymbol{i},-\boldsymbol{j}} = \boldsymbol{s}\right) = \rho\left(\boldsymbol{A}_{\boldsymbol{i},-\boldsymbol{j}} = \boldsymbol{s};\alpha\right)$ and $\rho\left(\boldsymbol{A}_{\boldsymbol{i},-\boldsymbol{j}} = \boldsymbol{s};\alpha\right)$ follows Definition 1(B) under type B parameterization.

eAppendix 2. Proof of Theorem 1 (overall effect partitioning)

$$OE(\alpha, \alpha') = \bar{Y}(\alpha) - \bar{Y}(\alpha')$$

$$= [\alpha \cdot \bar{Y}(1;\alpha) + (1-\alpha) \cdot \bar{Y}(0;\alpha)] - [\alpha' \cdot \bar{Y}(1;\alpha') + (1-\alpha') \cdot \bar{Y}(0;\alpha')]$$

$$= \alpha \cdot \bar{Y}(1;\alpha) + \bar{Y}(0;\alpha) - \alpha \cdot \bar{Y}(0;\alpha) - \alpha' \cdot \bar{Y}(1;\alpha') - \bar{Y}(0;\alpha') + \alpha' \cdot \bar{Y}(0;\alpha')$$

$$= \alpha \cdot \bar{Y}(1;\alpha) + \bar{Y}(0;\alpha) + [(\alpha' - \alpha) \cdot \bar{Y}(0;\alpha) - \alpha' \cdot \bar{Y}(0;\alpha)] - [(\alpha' - \alpha) \cdot \bar{Y}(1;\alpha') + \alpha \cdot \bar{Y}(1;\alpha')] - \bar{Y}(0;\alpha') + \alpha' \cdot \bar{Y}(0;\alpha')$$

$$= \alpha \cdot \bar{Y}(1;\alpha) + \bar{Y}(0;\alpha) + (\alpha' - \alpha) \cdot (\bar{Y}(0;\alpha) - \bar{Y}(1;\alpha')) - \alpha' \cdot \bar{Y}(0;\alpha) - \alpha \cdot \bar{Y}(1;\alpha') - \bar{Y}(0;\alpha') + \alpha' \cdot \bar{Y}(0;\alpha')$$

$$= \alpha \cdot \left(\bar{Y}(1;\alpha) - \bar{Y}(1;\alpha')\right) + (\alpha' - \alpha) \cdot \left(\bar{Y}(0;\alpha) - \bar{Y}(1;\alpha')\right) + (1-\alpha') \cdot \left(\bar{Y}(0;\alpha) - \bar{Y}(0;\alpha')\right)$$

$$= \alpha \cdot IE^{Vax}(\alpha, \alpha') + (\alpha' - \alpha) \cdot TE(\alpha, \alpha') + (1-\alpha') \cdot IE^{Unvax}(\alpha, \alpha').$$

The first line follows by definition; the second follows from line 1 of proof of Theorem 3 from VanderWeele and Tchetgen Tchetgen[1]; the third distributes terms, the fourth adds and subtracts $\alpha' \cdot \bar{Y}(0;\alpha)$ and $\alpha \cdot \bar{Y}(1;\alpha')$; the fifth and sixth rearrange terms; and the last applies definitions of IE^{Vax}, TE, and IE^{Unvax}. The proof can also be shown visually as illustrated in Figure 1.

eAppendix 3. The susceptible-infected-recovered-death (SIRD) model

1. Model structure

eFigure 1 shows the flowchart for the transmission model described in equation (5) of the main text.

$$
\begin{tikzcd}
S_{u,\alpha} \arrow[r, "\lambda_\alpha"] & I_{u,\alpha} \arrow[r, "(1-\mu)\cdot\gamma"] \arrow[dr, "\mu\cdot\gamma"'] & R_{u,\alpha} \\
 & & D_{u,\alpha} \\
S_{v,\alpha} \arrow[r, "\theta\cdot\lambda_\alpha"] & I_{v,\alpha} \arrow[r, "(1-\kappa\cdot\mu)\cdot\gamma"] \arrow[dr, "\kappa\cdot\mu\cdot\gamma"'] & R_{v,\alpha} \\
 & & D_{v,\alpha}
\end{tikzcd}
$$

eFigure 1. Model flowchart.

2. Model parameters

To verify Claim 1a, a pair of trajectories are simulated with same parameters except for the proportions vaccinated ($\alpha_0 = 0$ versus $\alpha_1 = 0.7$). Similarly, to verify Claim 1b, another pair of trajectories are simulated with $\alpha_1 = 0.7$ versus $\alpha_2 = 0.9$. Epidemic trajectories are simulated based on the SIRD model in equation (5)—first under Scenario 1 (i.e., time-invariant parameters) and then Scenarios 2 to 5 (i.e., time-dependent parameters). eTable 1 specifies parameter values for simulations. All models are implemented using R package *odin,*[3] whereas the actual solution of the differential equations is conducted with the *deSolve* package using numerical solvers "lsoda" and "ode45."[4]

eTable 1. List of parameters for simulations

Parameter	Baseline value [a]	Definition	Variation over time (if any)

$\boldsymbol{\beta}$	0.15	The number of effective contacts made by a typical infectious individual per day	Scenario 2: β increases to 0.6 from Day 300 onwards
$\boldsymbol{\mu}$	0.01	Probability of death due to infection	Scenario 3: μ increases to 0.1 from Day 300 onwards
$\boldsymbol{\theta}$	0.5	1 – vaccine efficacy against infection ($VE_{infection}$)/100%	Scenario 4: $VE_{infection}$ wanes linearly after Day 100, reaching 0% at Day 300
$\boldsymbol{\kappa}$	0.1	1 – vaccine efficacy against death given infection ($VE_{death\|infection}$)/100%	Scenario 4: $VE_{death\|infection}$ wane linearly after Day 100, reaching 0% at Day 300
$\boldsymbol{\gamma}$	0.07	Recovery rate per day	-

[a] All parameters are time-invariant under Scenario 1. Scenarios 2 to 5 change one or two parameters separately while holding other parameters at the baseline values.

eTable 2. List of initial conditions in a group wherein the proportion vaccinated is α

Variable	Initial condition(s)	Definition
$\boldsymbol{N(0)}$	20,000	Number alive at baseline
$\boldsymbol{S_{u,\alpha}(0)}$	$(20{,}000 - 20) * (1 - \alpha)$	Number unvaccinated susceptible at baseline
$\boldsymbol{S_{v,\alpha}(0)}$	$(20{,}000 - 20) * \alpha$	Number vaccinated susceptible at baseline
$\boldsymbol{I_{u,\alpha}(0)}$	$20 * (1 - \alpha)$	Number unvaccinated infectious at baseline
$\boldsymbol{I_{v,\alpha}(0)}$	$20 * \alpha$	Number vaccinated infectious at baseline
$\boldsymbol{R_{u,\alpha}(0)}$	0	Number unvaccinated recovered at baseline
$\boldsymbol{R_{v,\alpha}(0)}$	0	Number vaccinated recovered at baseline

$\boldsymbol{D_{u,\alpha}(0)}$	0	Number unvaccinated who died due to infection at baseline
$\boldsymbol{D_{v,\alpha}(0)}$	0	Number vaccinated who died due to infection at baseline

eAppendix 4. Proving Claim 1 under Scenario 1 in the SIRD model at $t \to \infty$

1. Definitions

Let $0 \leq \alpha_0 < \alpha_1 < \alpha_2 \leq 1$. Recall Claim 1:

$$\delta^O(t, \alpha_0, \alpha_1) \geq \delta^D(t, \alpha_0, \alpha_1) \qquad \forall\, \alpha_0 < \alpha_1,$$

and

$$\delta^O(t, \alpha_1, \alpha_2) \geq \delta^D(t, \alpha_2, \alpha_1) \qquad \forall\, \alpha_2 > \alpha_1.$$

By expanding equations (4.1) and (4.2) and suppressing the time notation (t), we have:

$$\begin{aligned}\delta^O(\alpha_0, \alpha_1) &= N \cdot \alpha_0 \cdot IE^{Vax}(\alpha_0, \alpha_1) + \delta^D(\alpha_0, \alpha_1) + N \cdot (1 - \alpha_0) \cdot IE^{Unvax}(\alpha_0, \alpha_1) \\ &= N \cdot \alpha_0 \cdot \left(\bar{Y}(1; \alpha_0) - \bar{Y}(1; \alpha_1)\right) + N \cdot (\alpha_1 - \alpha_0) \cdot \left(\bar{Y}(0; \alpha_1) - \bar{Y}(1; \alpha_1)\right) + N \\ &\quad \cdot (1 - \alpha_0) \cdot \left(\bar{Y}(0; \alpha_0) - \bar{Y}(0; \alpha_1)\right) \qquad (6.1)\end{aligned}$$

for $\alpha_1 > \alpha_0$, and

$$\begin{aligned}\delta^O(\alpha_1, \alpha_2) &= N \cdot \alpha_2 \cdot IIE^{Vax}(\alpha_1, \alpha_2) + \delta^D(\alpha_2, \alpha_1) + N \cdot (1 - \alpha_2) \cdot IE^{Unvax}(\alpha_1, \alpha_2) \\ &= N \cdot \alpha_2 \cdot \left(\bar{Y}(1; \alpha_1) - \bar{Y}(1; \alpha_2)\right) + N \cdot (\alpha_2 - \alpha_1) \cdot \left(\bar{Y}(0; \alpha_1) - \bar{Y}(1; \alpha_1)\right) + N \\ &\quad \cdot (1 - \alpha_2) \cdot \left(\bar{Y}(0; \alpha_1) - \bar{Y}(0; \alpha_2)\right) \qquad (6.2)\end{aligned}$$

for $\alpha_2 > \alpha_1$.

We first give an intuitive explanation of $\delta^D(\alpha_0, \alpha_1)$ being a lower bound of $\delta^O(\alpha_0, \alpha_1)$. Recall that individuals fall into three categories based on vaccination status under the counterfactuals with α_0 or α_1 proportion vaccinated: 1) Those who are unvaccinated under both counterfactuals (referred to as "never-vaccinated" and represented by the dotted region in Figure 1); 2) those who are unvaccinated under α_0 but vaccinated under α_1 (referred to as "additionally-vaccinated" and represented by the gridded region in Figure 1); and 3) those who are vaccinated under both counterfactuals (referred to as "always-vaccinated" and represented by the stripped region in Figure 1). If we update risk of the "always-vaccinated" under α_0 from $\bar{Y}(1; \alpha_0)$ to $\bar{Y}(1; \alpha_1)$ and

update risk of the "never-vaccinated" under α_0 from $\bar{Y}(0;\alpha_0)$ to $\bar{Y}(0;\alpha_1)$, then we have an updated right-hand side of equation (6.1):

$$N \cdot \alpha_0 \cdot \left(\bar{Y}(1;\alpha_1) - \bar{Y}(1;\alpha_1)\right) + N \cdot (\alpha_1 - \alpha_0) \cdot \left(\bar{Y}(0;\alpha_1) - \bar{Y}(1;\alpha_1)\right) + N \cdot (1 - \alpha_1) \cdot \left(\bar{Y}(0;\alpha_1) - \bar{Y}(0;\alpha_1)\right) = N \cdot (\alpha_1 - \alpha_0) \cdot \left(\bar{Y}(0;\alpha_1) - \bar{Y}(1;\alpha_1)\right) = \delta^D(t, \alpha_0, \alpha_1)$$

which is reduced to $\delta^D(\alpha_0, \alpha_1)$ and is lower than the actual overall effect because the original risk $\bar{Y}(1;\alpha_0)$ in the first term of the last line of equation (6.1) is higher than the updated $\bar{Y}(1;\alpha_1)$ (to be proved in Section 3), and the original risk $\bar{Y}(0;\alpha_0)$ in the last term of the last line of equation (6.1) is also higher than the updated $\bar{Y}(0;\alpha_1)$ (to be proved in Section 3). Therefore, $\delta^D(\alpha_0, \alpha_1)$ in the updated equation is lower than $\delta^O(\alpha_0, \alpha_1)$ in the original equation (6.1) and Claim 1 holds for $\alpha_1 > \alpha_0$.

Similarly, for $\alpha_2 > \alpha_1$, if we update risk of "always-vaccinated" under α_2 from $\bar{Y}(1;\alpha_2)$ to $\bar{Y}(1;\alpha_1)$ and update risk of "never-vaccinated" under α_2 from $\bar{Y}(0;\alpha_2)$ to $\bar{Y}(0;\alpha_1)$, we have on the right-hand side of equation (6.2):

$$N \cdot \alpha_2 \cdot \left(\bar{Y}(1;\alpha_1) - \bar{Y}(1;\alpha_1)\right) + N \cdot (\alpha_2 - \alpha_1) \cdot \left(\bar{Y}(0;\alpha_1) - \bar{Y}(1;\alpha_1)\right) + N \cdot (1 - \alpha_2) \cdot \left(\bar{Y}(0;\alpha_1) - \bar{Y}(0;\alpha_1)\right) = (\alpha_2 - \alpha_1) \cdot \left(\bar{Y}(0;\alpha_1) - \bar{Y}(1;\alpha_1)\right) = \delta^D(t, \alpha_2, \alpha_1)$$

which is reduced to $\delta^D(\alpha_2, \alpha_1)$ and is lower than the actual overall effect because the original risk $\bar{Y}(1;\alpha_2)$ in the first term of the last line of equation (6.2) is lower than the updated $\bar{Y}(1;\alpha_1)$ (to be proved in Section 3), and the original risk $\bar{Y}(0;\alpha_2)$ in the last term is also lower than the updated $\bar{Y}(0;\alpha_1)$ (to be proved in Section 3). Therefore, $\delta^D(\alpha_2, \alpha_1)$ in the updated equation is lower than $\delta^O(\alpha_1, \alpha_2)$ in the original equation (6.2) and Claim 1 holds for $\alpha_2 > \alpha_1$. Next, we will formally prove Claim 1 under Scenario 1 at $t \to \infty$.

2. Sufficient Conditions for Claim 1 to hold

Based on equation (4.1), Claim 1 holds for all $\alpha_0 < \alpha_1$ when $IE^{Vax}(t, \alpha_0, \alpha_1)$ and $IE^{Unvax}(t, \alpha_0, \alpha_1)$ are both non-negative. Similarly, based on equation (4.2), Claim 1 holds for all $\alpha_2 > \alpha_1$ when $IE^{Vax}(t, \alpha_1, \alpha_2)$ and $IE^{Unvax}(t, \alpha_1, \alpha_2)$ are both non-negative.

3. Proof of Claim 1 under Scenario 1 in the Susceptible-infected-recovered (SIR) Model at $t \to \infty$

When probability of death due to infection is zero, the SIRD model in equation (5) reduces to a SIR model stratified by vaccination status. $IE^{Unvax}(\infty, \alpha_0, \alpha_1)$ is positive for $\alpha_1 > \alpha_0$ in the SIR model by proving that $\bar{Y}(\infty, 0; \alpha)$, the final epidemic fraction in the unvaccinated individuals (i.e., the fraction of the unvaccinated individuals that became infected at the end of the outbreak), decreases with α. Similarly, $IE^{Vax}(\infty, \alpha_0, \alpha_1)$ is positive by proving that $\bar{Y}(\infty, 1; \alpha)$, the final epidemic fraction in the vaccinated individuals, also decreases with α.

3.1 Proof that $\bar{Y}(\infty, 0; \alpha)$ and $\bar{Y}(\infty, 1; \alpha)$ decrease with α

To prove that $\bar{Y}(\infty, 0; \alpha)$ and $\bar{Y}(\infty, 1; \alpha)$ decrease with the proportion vaccinated α, we prove that $\bar{Y}(\infty; \alpha)$, the average final epidemic fraction, decreases with α. $\bar{Y}(\infty; \alpha)$ is the weighted average of $\bar{Y}(\infty, 0; \alpha)$ and $\bar{Y}(\infty, 1; \alpha)$.[5] Assuming that 1) the initial proportion susceptible is close to one, 2) the total group size N is near infinite, and 3) transmission events follow a Poisson process, the average final epidemic size is:

$$\bar{Y}(\infty; \alpha) = (1 - \alpha) \cdot \bar{Y}(\infty, 0; \alpha) + \alpha \cdot \bar{Y}(\infty, 1; \alpha)$$

$$= (1 - \alpha) \cdot \left(1 - exp\left(-R_0 \bar{Y}(\infty; \alpha)\right)\right) + \alpha \left(1 - exp\left(-\theta R_0 \bar{Y}(\infty; \alpha)\right)\right)$$

$$= \left(1 - exp\left(-R_0 \bar{Y}(\infty; \alpha)\right)\right) - \alpha \left(exp\left(-\theta R_0 \bar{Y}(\infty; \alpha)\right) - exp\left(-R_0 \bar{Y}(\infty; \alpha)\right)\right) \quad (7)$$

where $\theta = 1 - VE_{infection}/100\%$ given a leaky vaccine, and R_0 is the basic reproduction number, defined as the expected number of new infections caused by a single infected individual during

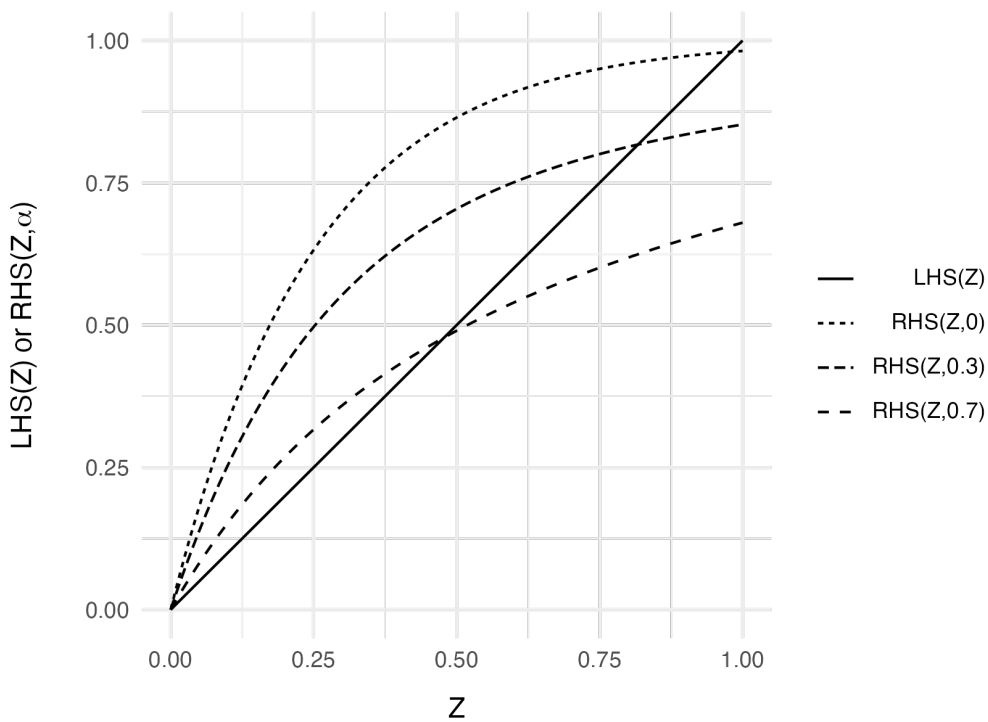
the infectious period in a completely susceptible population. Lin et al. (2024) arrived at the same final epidemic size equation by also considering a leaky vaccine in a stratified susceptible-infected-recovered model. See equation S20 of Lin et al.[6]

Define two separate functions with independent arguments, Z and α (Miller[5] used a similar type of argument to arrive at the final epidemic size):

$$\mathrm{LHS}(Z,\alpha) = \mathrm{LHS}(Z) = Z, \tag{8}$$

$$\mathrm{RHS}(Z,\alpha) = \big(1 - exp(-R_0 Z)\big) - \alpha\big(exp(-\theta R_0 Z) - exp(-R_0 Z)\big). \tag{9}$$

Here the equality $\mathrm{LHS}(\cdot) = \mathrm{RHS}(\cdot)$ is not assumed; in particular, $\mathrm{RHS}(\cdot)$ takes two independent arguments, Z and α. Plot $\mathrm{LHS}(Z)$ and $\mathrm{RHS}(Z,\alpha)$ in one graph, where $R_0 = 4$, $\theta = 0.2$, and $\alpha = 0$, 0.3 and 0.7.



$\mathrm{LHS}(Z) = \mathrm{RHS}(Z,\alpha)$ when $Z = 0$, which is a trivial solution. We will now prove that if a positive (non-trivial) root exists such that $\mathrm{LHS}(Z) = \mathrm{RHS}(Z,\alpha)$, the value of this root decreases when α increases.

We first show that the first derivative of $\mathrm{RHS}(Z,\alpha)$ w.r.t Z at zero is greater than 1, which is the derivative of $\mathrm{LHS}(Z)$. That is:

$$\frac{d\mathrm{RHS}(Z,\alpha)}{dZ} = (1-\alpha)R_0 exp(-R_0 Z) + \alpha\theta R_0 \exp(-\theta R_0 Z)$$

$$\frac{d\mathrm{RHS}(Z,\alpha)}{dZ}|_{Z=0} = (1-\alpha)R_0 + \alpha\theta R_0 = R_e > 1 = \frac{d\mathrm{LHS}(Z)}{dZ}|_{Z=0}$$

where R_e is the effective reproduction number when proportion α of the group is vaccinated with a leaky vaccine.[7] The above inequality holds because R_e must be greater than 1 in order for any epidemic to occur, meaning that for some $\epsilon > 0$, $\mathrm{RHS}(\epsilon,\alpha) > LHS(\epsilon)$. In other words, $\mathrm{RHS}(Z,\alpha)$ exceeds $\mathrm{LHS}(Z)$ for small values of Z.

Next, we prove that $\mathrm{RHS}(Z,\alpha)$ is a concave function by computing the second derivative w.r.t Z:

$$\frac{d^2\mathrm{RHS}(Z,\alpha)}{dZ^2} = (\alpha-1)R_0^2 exp(-R_0 Z) - \alpha\theta^2 R_0^2 \exp(-\theta R_0 Z) < 0$$

for $\alpha \in [0,1]$. Because the second derivative is negative for $\alpha \in [0,1]$, $\mathrm{RHS}(Z,\alpha)$ is a concave function of Z for any possible values of α. Taken together, the first and second derivatives indicate that $\mathrm{RHS}(Z,\alpha)$ is above $\mathrm{LHS}(Z)$ when Z is greater than but close to zero, and $\mathrm{RHS}(Z,\alpha)$ bends towards $\mathrm{LHS}(Z)$ (i.e., the diagonal) and by assumption (that there is a nontrivial root of $\mathrm{LHS}(Z) = \mathrm{RHS}(Z,\alpha)$) coincides with the diagonal at some value of Z (See plot above).

Finally, consider two values $\alpha_1 > \alpha_0$ giving rise to two curves $\mathrm{RHS}(Z,\alpha=\alpha_0)$ and $\mathrm{RHS}(Z,\alpha=\alpha_1)$. Let $\alpha_1 = 0.7$ and $\alpha_0 = 0.3$ as the plot illustrates. Because $0 < \theta < 1$, $[exp(-\theta R_0 Z) - exp(-R_0 Z)]$ in equation (9) is always positive at any given $Z > 0$. Therefore, $\mathrm{RHS}(Z,\alpha=\alpha_0) > \mathrm{RHS}(Z,\alpha=\alpha_1)$ at any given $Z > 0$ when $\alpha_1 > \alpha_0$. Suppose the upper curve $\mathrm{RHS}(Z,\alpha_0)$ meets $\mathrm{LHS}(Z)$ at $Z = z_0$. Then $\mathrm{RHS}(z_0,\alpha_1) < z_0 = \mathrm{LHS}(z_0)$. Given that all expressions are continuous and that $\mathrm{RHS}(\epsilon,\alpha_0) > \mathrm{LHS}(\epsilon)$, there must have been some value z_1 to

the left of z_0 where $\mathrm{RHS}(z_1, \alpha_1)$ crossed $\mathrm{LHS}(z_1)$. In other words, $\mathrm{RHS}(z_1, \alpha_1) = \mathrm{LHS}(z_1)$ for some value $z_1 < z_0$, proving that the root z_1 of $\mathrm{RHS}(z_1, \alpha_1) = \mathrm{LHS}(z_1)$ is less than the root z_0 of $\mathrm{RHS}(z_0, \alpha_0) = \mathrm{LHS}(z_0)$ for $\alpha_1 > \alpha_0$, QED. These non-trivial roots are unique because $\mathrm{RHS}(Z, \alpha)$ is concave.

Any value Z that satisfies $\mathrm{LHS}(Z) = \mathrm{RHS}(Z, \alpha)$ will satisfy equation (7). Suppose z_0 is a positive root for $\mathrm{LHS}(Z) = \mathrm{RHS}(Z, \alpha = \alpha_0)$ such that $\pi_{\alpha_0}(\infty) = z_0$, and similarly suppose z_1 is a positive root for $\mathrm{LHS}(Z) = \mathrm{RHS}(Z, \alpha = \alpha_1)$ such that $\pi_{\alpha_1}(\infty) = z_1$. Therefore, we have $\pi_{\alpha_1}(\infty) < \pi_{\alpha_0}(\infty)$ when $\alpha_1 > \alpha_0$.

Given that $\bar{Y}(\infty, 0; \alpha) = 1 - \exp\left(-R_0 \bar{Y}(\infty; \alpha)\right)$,[5] $\bar{Y}(\infty, 0; \alpha)$ decreases with α because $\bar{Y}(\infty; \alpha)$ decreases with α. Similarly, given that $\bar{Y}(\infty, 1; \alpha) = 1 - exp\left(-\theta R_0 \bar{Y}(\infty; \alpha)\right)$,[5] $\bar{Y}(\infty, 1; \alpha)$ decreases with α.

3.2 Proof of $IE^{Unvax}(\infty, \alpha_0, \alpha_1) > 0$ and $IE^{Vax}(\infty, \alpha_0, \alpha_1) > 0$ for $\alpha_1 > \alpha_0$

In Section 3.1, we proved that $\bar{Y}(\infty, 0; \alpha)$ decreases with α. Therefore, $IE^{Unvax}(\infty, \alpha_0, \alpha_1) = \bar{Y}(\infty, 0; \alpha_0) - \bar{Y}(\infty, 0; \alpha_1) > 0$ for $\alpha_1 > \alpha_0$. We also proved that $\pi_{1,\alpha}(\infty)$ decreases with α, and therefore $IIE^{Vax}(\infty, \alpha_0, \alpha_1) = \bar{Y}(\infty, 1; \alpha_0) - \bar{Y}(\infty, 1; \alpha_1) > 0$ for $\alpha_1 > \alpha_0$.

Since $IE^{Unvax}(\infty, \alpha_0, \alpha_1)$ and $IE^{Vax}(\infty, \alpha_0, \alpha_1)$ are positive, Claim 1 holds for any $\alpha_1 > \alpha_0$ at $t \to \infty$. The proof in Section 3 also applies to any $\alpha_2 > \alpha_1$ in showing that $IIE^{Unvax}(\infty, \alpha_1, \alpha_2)$ and $IIE^{Vax}(\infty, \alpha_1, \alpha_2)$ are positive, such that Claim 1 holds under Scenario 1 in the Susceptible-infected-recovered (SIR) Model at $t \to \infty$.

4. Proof of Claim 1 under Scenario 1 in the Susceptible-infected-recovered-death (SIRD) Model at $t \to \infty$

Now consider the SIRD model with non-zero probability of death due to infection. Based on the transmission dynamics in equation (5), those who died of infection are fixed proportion of those who became infected at the end of epidemic. We have:

$$\bar{Y}^{death}(\infty, 0; \alpha) = \mu \cdot \bar{Y}^{infection}(\infty, 0; \alpha) \quad (10.1)$$

for the unvaccinated individuals and

$$\bar{Y}^{death}(\infty, 1; \alpha) = \kappa \cdot \mu \cdot \bar{Y}^{infection}(\infty, 1; \alpha) \quad (10.2)$$

for the vaccinated individuals.

For the unvaccinated individuals, by substituting the right-hand side of equation (10.1) into the equation of $IE^{Unvax,death}$, we have:

$$\begin{aligned} IE^{Unvax,death}(\infty, \alpha_0, \alpha_1) &= \bar{Y}^{death}(\infty, 0; \alpha_0) - \bar{Y}^{death}(\infty, 0; \alpha_1) \\ &= \mu \cdot \left(\bar{Y}^{infection}(\infty, 0; \alpha_0) - \bar{Y}^{infection}(\infty, 0; \alpha_1)\right) \\ &= \mu \cdot IE^{Unvax,infection}(\infty, \alpha_0, \alpha_1). \end{aligned}$$

We show that $IE^{Unvax,death}(\infty, \alpha_0, \alpha_1)$ has the same direction as $IE^{Unvax,infection}(\infty, \alpha_0, \alpha_1)$. If $IE^{Unvax,infection}(\infty, \alpha_0, \alpha_1)$ is positive (Section 3), so is $IE^{Unvax,death}(\infty, \alpha_0, \alpha_1)$.

Similarly, for the vaccinated individuals, by substituting the right-hand side of equation (10.2) into the equation of $IE^{Vax,death}$, we have:

$$\begin{aligned} IE^{Vax,death}(\infty, \alpha_0, \alpha_1) &= \bar{Y}^{death}(\infty, 1; \alpha_0) - \bar{Y}^{death}(\infty, 1; \alpha_1) \\ &= \kappa \cdot \mu \cdot \left(\bar{Y}^{infection}(\infty, 1; \alpha_0) - \bar{Y}^{infection}(\infty, 1; \alpha_1)\right) \\ &= \kappa \cdot \mu \cdot IE^{Vax,infection}(\infty, \alpha_0, \alpha_1). \end{aligned}$$

If $IE^{Vax,infection}(\infty, \alpha_0, \alpha_1)$ is positive (Section 3), so is $IE^{Vax,death}(\infty, \alpha_0, \alpha_1)$.

In sum, $IE^{Vax}(\infty, \alpha_0, \alpha_1)$ and $IE^{Unvax}(\infty, \alpha_0, \alpha_1)$ are positive for both death and infection given $\alpha_1 > \alpha_0$ (similarly, $IE^{Vax}(\infty, \alpha_1, \alpha_2)$ and $IE^{Unvax}(\infty, \alpha_1, \alpha_2)$ are positive for $\alpha_2 > \alpha_1$). Therefore, Claim 1 holds under Scenario 1 in the Susceptible-infected-recovered-death (SIRD) model at $t \to \infty$.

eAppendix 5. Latin hypercube sampling to verify Claim 1 under Scenario 1 and Claim 1a under Scenario 4

1. Verifying Claim 1 under Scenario 1

As Figures 2 and 3 show, Claims 1a and 1b hold under Scenario 1 given the particular proportions vaccinated ($\alpha_0 = 0$ versus $\alpha_1 = 0.7$ for Claim 1a and $\alpha_1 = 0.7$ versus $\alpha_2 = 0.9$ for Claim 1b) and parameter values specified in eTable 1. To test the robustness of Claim 1 under alternative parameter combinations, Latin hypercube sampling (LHS) is used to randomly draw 1,000 sets of proportions vaccinated and parameters. Initial conditions are same as in eTable 2. LHS samples are first generated from uniform distributions using the *lhs* package[8] and then scaled to match the range of each parameter as shown in eTable 3.

eTable 3. Parameter ranges for Latin hypercube sampling under Scenarios 1 and 4

Parameters for both Scenarios 1 and 4		
Parameter	**Range**	**Definition**
$\boldsymbol{R_0}$	[1, 20]	Basic reproduction number ($R_0 = \beta/\gamma$)
$\boldsymbol{\alpha_1}$	[0.01, 1]	Proportion vaccinated at baseline for one group
$\boldsymbol{\alpha_0}$ or $\boldsymbol{\alpha_2}$ [a]	[0.01, 1]	Proportion vaccinated at baseline for the comparator group under Claim 1
$\boldsymbol{\mu}$	[0, 0.1]	Probability of death due to infection
$\boldsymbol{\gamma}$	[1/3, 1/14]	Recovery rate per day
$\boldsymbol{\theta}$	[0, 1]	$1 - VE_{infection}$/100% at baseline
$\boldsymbol{\kappa}$	[0, 1]	$1 - VE_{death\|infection}$/100% at baseline
Additional parameters for Scenario 4		

$\boldsymbol{T_{infection,unwaned}}$	[0, 365]	Time (days since baseline) after which $VE_{infection}$ starts to wane
$\boldsymbol{T_{infection,waning}}$	[0, 365]	Duration for $VE_{infection}$ to wane to 0%
$\boldsymbol{T_{death,unwaned}}$	[0, 365]	Time (days since baseline) after which $VE_{death\|infection}$ starts to wane
$\boldsymbol{T_{death,waning}}$	[0, 365]	Duration for $VE_{death\|infection}$ to wane to 0%

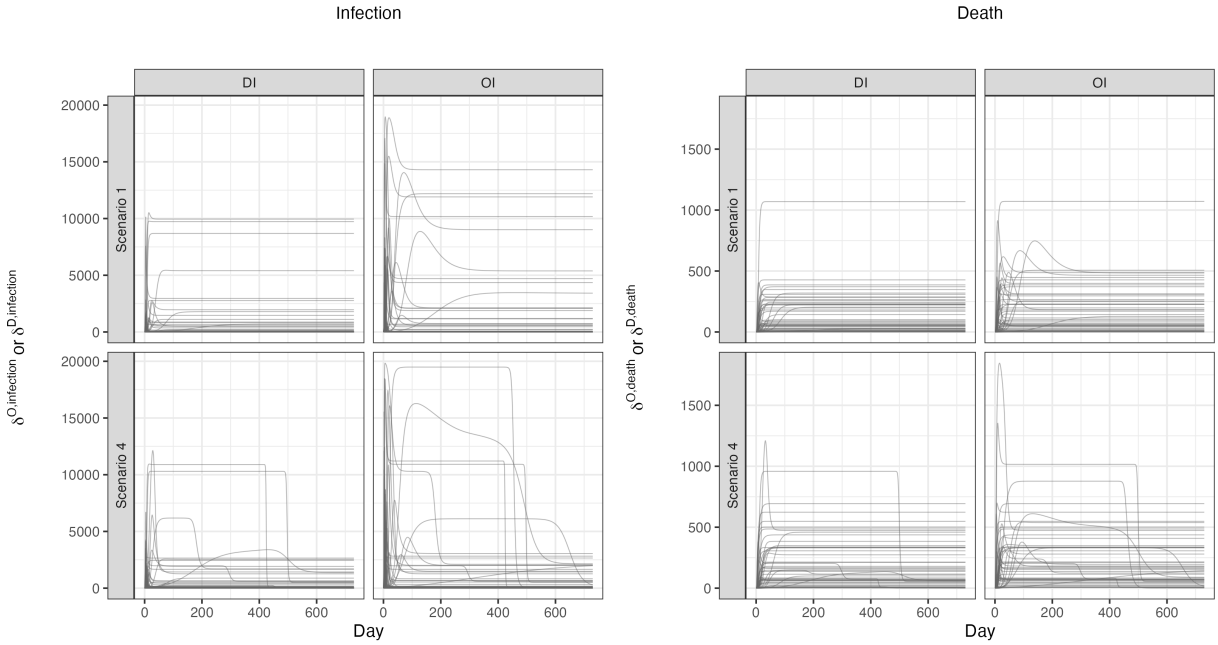
$\boldsymbol{VE_{infection}}$ is vaccine efficacy against infection; $\boldsymbol{VE_{death|infection}}$ is vaccine efficacy against death given infection.

[a] $\boldsymbol{\alpha_0}$ or $\boldsymbol{\alpha_2}$ is randomly drawn to verify Claim 1 only; $\boldsymbol{\alpha_0 = 0}$ for Claim 1a by definition.

To exclude results due to roundoff errors during numerical integration of the differential equations, Claim 1 is disproven based on the criteria $\delta^O - \delta^D < -10^{-7}$ at any time t, rather than the conventional criteria of < 0. From the LHS procedure, Claim 1 holds under Scenario 1 for all the 1,000 iterations. eFigure 2 shows direct impact and overall impact for 50 random LHS samples.

2. Verifying Claim 1a under Scenario 4

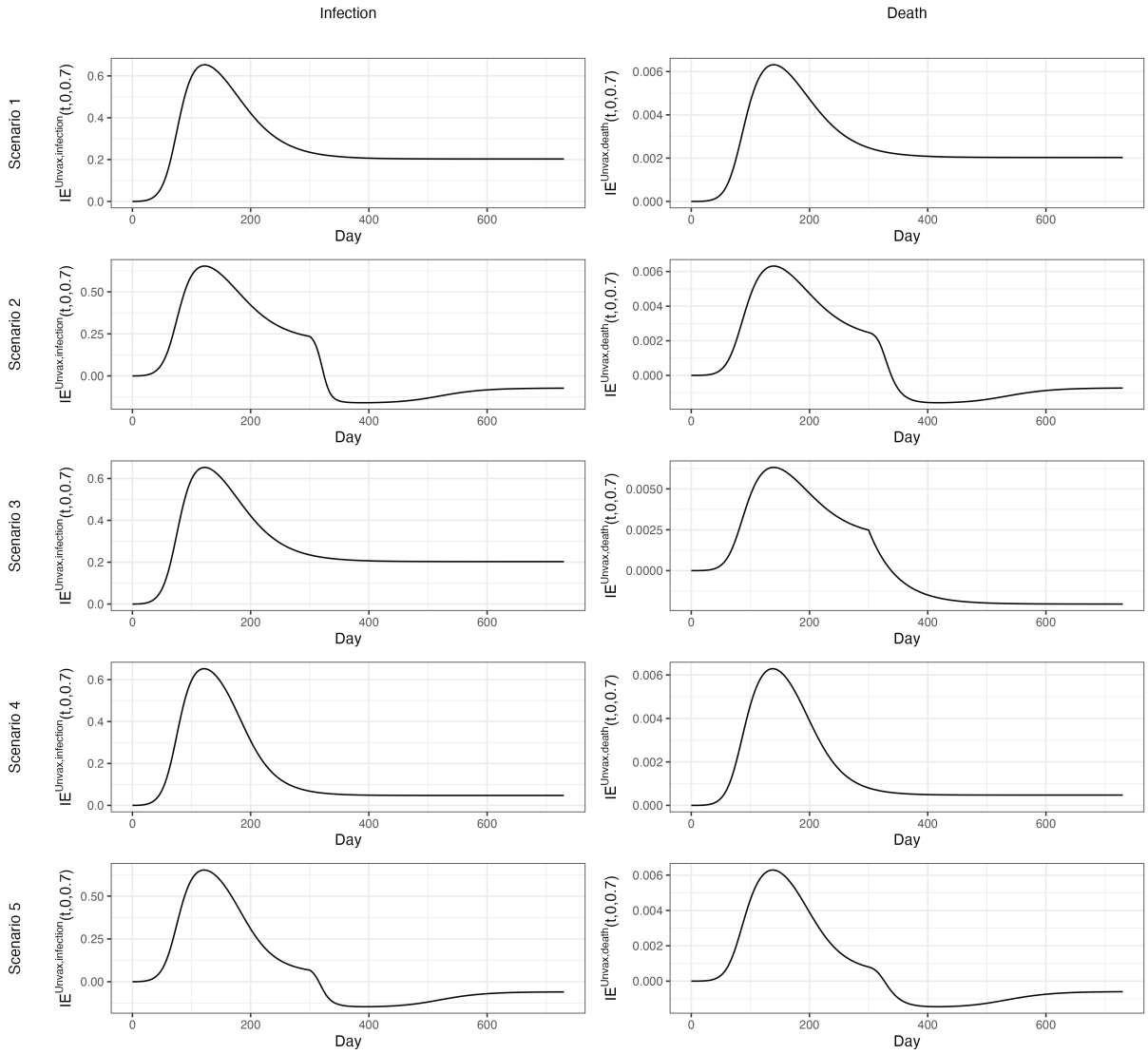
Figure 2 also shows that Claim 1a holds under Scenario 4 given the parameter values specified in eTable 1. Similarly, LHS is conducted to verify if Claim 1a holds under Scenario 4 with different values for α_1 (while holding constant $\alpha_0 = 0$) and alternative sets of model parameters; parameter ranges are specified in eTable 3. From the LHS procedure, Claim 1a holds under Scenario 4 for all the 1,000 iterations (eFigure 2).



eFigure 2. Direct impact and overall impact to verify Claim 1 under Scenario 1 and to verify Claim 1a under Scenario 4 from 50 Latin hypercube samples.

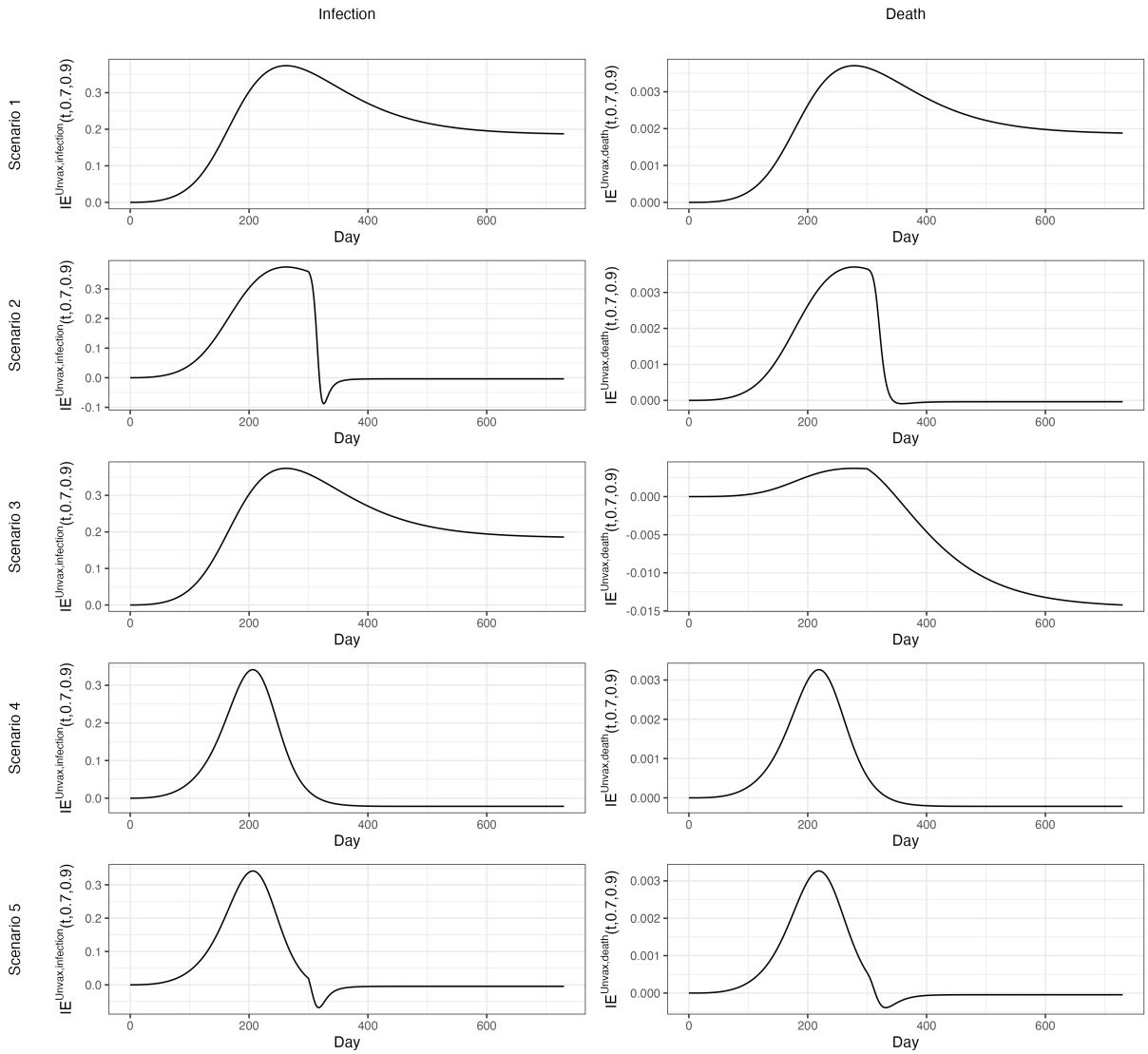
eAppendix 6. Trajectories of indirect effects for the vaccinated and unvaccinated.

Based on equation (4.1), Claim 1a does not hold when $IE^{Unvax}(t, 0, 0.7)$ is negative, as shown in eFigure 3. Similarly, based on equation (4.2), Claim 1b may not hold when $IE^{Unvax}(t, 0.7, 0.9)$ or $IE^{Vax}(t, 0.7, 0.9)$ is negative, as shown in eFigures 4 and 5, respectively.

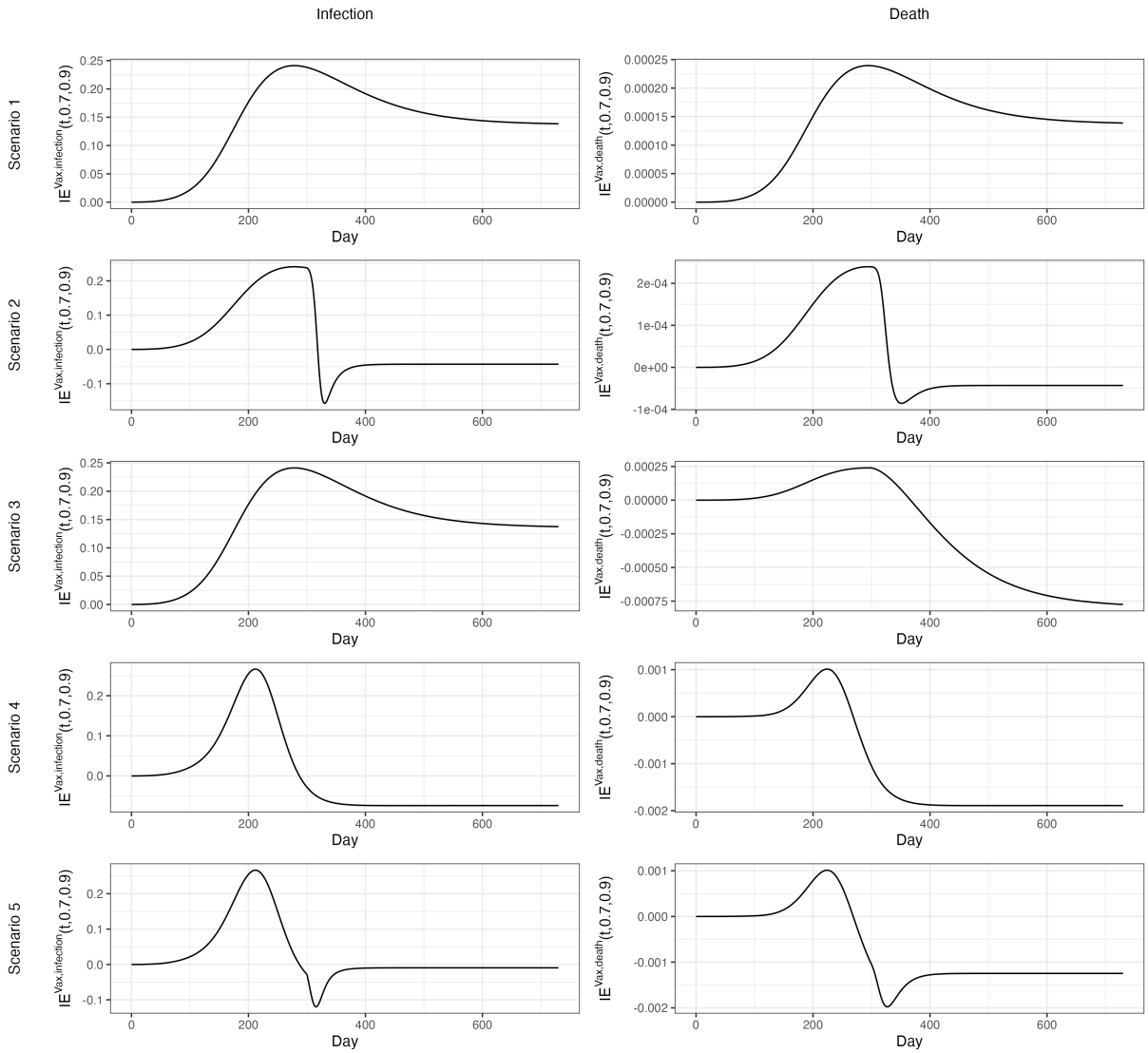


eFigure 3. Indirect effect for the unvaccinated $IE^{Unvax}(t, 0, 0.7)$. Scenario 1, all parameters are time-invariant; Scenario 2, the number of effective contacts made by a typical infectious

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.

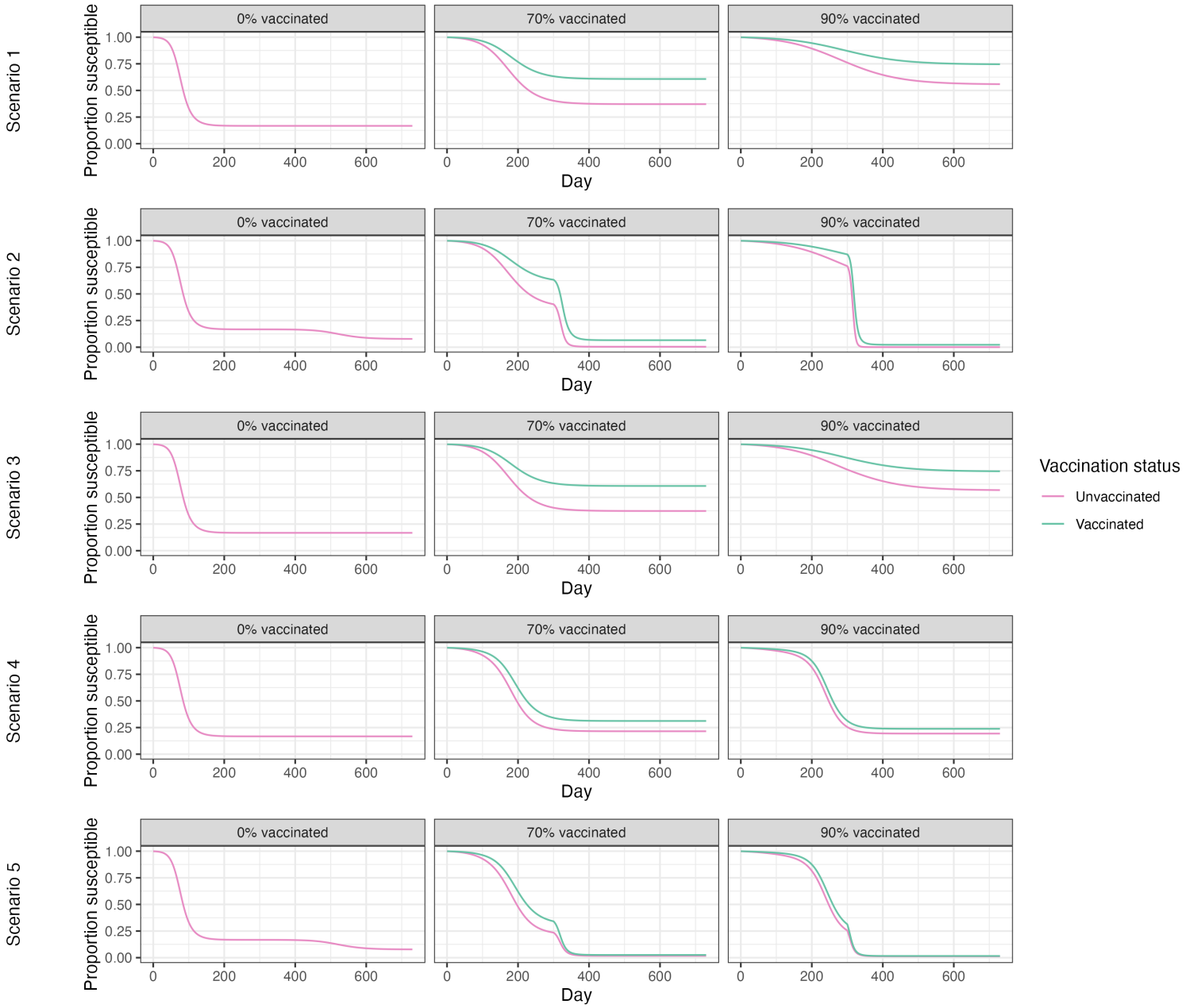


eFigure 4. Indirect effect for the unvaccinated $IE^{Unvax}(t, 0.7, 0.9)$.



eFigure 5. Indirect effect for the vaccinated $IE^{Vax}(t, 0.7, 0.9)$.

eAppendix 7. Trajectories of proportion susceptible by vaccination status over time

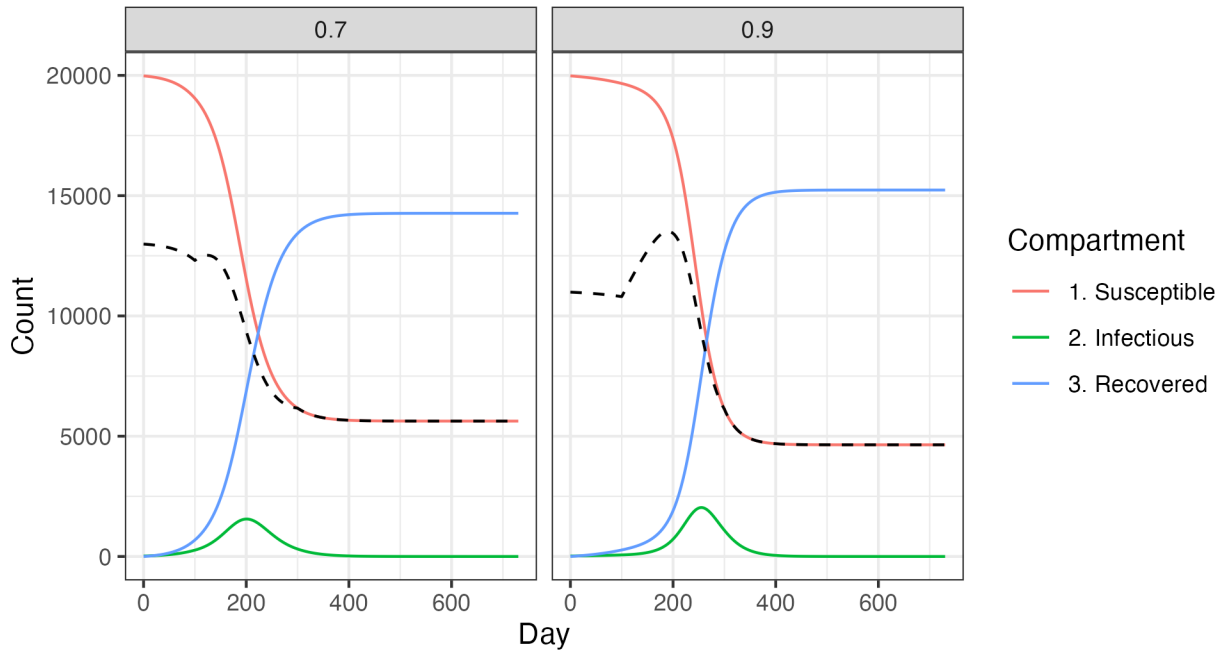


eFigure 6. Proportion susceptible among vaccinated and unvaccinated individuals under 0%, 70%, or 90% proportions vaccinated. Scenario 1, all parameters are time-invariant; Scenario 2, the number of effective contacts made by a typical infectious individual per day (β) increases from 0.15 to 0.6 at Day 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.

eAppendix 8. Epidemic curves under Scenario 4.

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.



eFigure 7. Epidemic curves under Scenario 4 given $\alpha_1 = 0.7$ and $\alpha_2 = 0.9$ proportions vaccinated. The dashed line represents the number of effectively susceptible persons $= S_u(t) + \theta(t) \cdot S_v(t)$ where $S_u(t)$ is the number of unvaccinated susceptibles and $\theta(t) \cdot S_v(t)$ is the number of vaccinated susceptibles multiplied with $\theta(t) = 1 - VE_{infection}(t)/100\%$. The dashed lines show that more individuals at risk of infection (weighted by that risk) are built up under $\alpha_2 = 0.9$ than $\alpha_1 = 0.7$ before the epidemic peaks.

References for eAppendix

1. 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

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. FitzJohn R. odin: ODE Generation and Integration. Published online 2024. https://github.com/mrc-ide/odin

4. Soetaert K, Petzoldt T, R. Woodrow S. Solving Differential Equations in R: Package deSolve. *Journal of Statistical Software*. 2010;33(9):1-25. doi:10.18637/jss.v033.i09

5. Miller JC. A Note on the Derivation of Epidemic Final Sizes. *Bull Math Biol*. 2012;74(9):2125-2141. doi:10.1007/s11538-012-9749-6

6. Lin L, Hamedmoghadam H, Shorten R, Stone L. Quantifying indirect and direct vaccination effects arising in the SIR model. *J R Soc Interface*. 2024;21(218):20240299. doi:10.1098/rsif.2024.0299

7. Chao DL, Dimitrov DT. Seasonality and the effectiveness of mass vaccination. *Mathematical Biosciences and Engineering*. 2016;13(2):249-259. doi:10.3934/mbe.2015001

8. Carnell R. lhs: Latin Hypercube Samples. Published online 2022. https://CRAN.R-project.org/package=lhs