

# UC Berkeley

## UC Berkeley Previously Published Works

### Title

Uses of pathogen detection data to estimate vaccine direct effects in case-control studies.

### Permalink

<https://escholarship.org/uc/item/3d82d0br>

### Journal

Journal of The Royal Society Interface, 17(169)

### ISSN

1742-5689

### Author

Lewnard, Joseph A

### Publication Date

2020-08-01

### DOI

10.1098/rsif.2020.0161

Peer reviewed

## Research



**Cite this article:** Lewnard JA. 2020 Uses of pathogen detection data to estimate vaccine direct effects in case–control studies. *J. R. Soc. Interface* **17**: 20200161.

<http://dx.doi.org/10.1098/rsif.2020.0161>

Received: 8 March 2020

Accepted: 16 July 2020

**Subject Category:**

Life Sciences–Mathematics interface

**Subject Areas:**

biomathematics

**Keywords:**

infectious disease, epidemiology, estimation, causal inference, vaccine effectiveness

**Author for correspondence:**

Joseph A. Lewnard

e-mail: [jlewnard@berkeley.edu](mailto:jlewnard@berkeley.edu)

# Uses of pathogen detection data to estimate vaccine direct effects in case–control studies

Joseph A. Lewnard<sup>1,2,3</sup>

<sup>1</sup>Division of Epidemiology, School of Public Health, <sup>2</sup>Division of Infectious Diseases and Vaccinology, School of Public Health, and <sup>3</sup>Center for Computational Biology, College of Engineering, University of California, Berkeley, CA 94720, USA

JAL, 0000-0002-8505-8839

The fact that many pathogens can be carried or shed without causing symptoms complicates the interpretation of microbiological data when diagnosing certain infectious disease syndromes. Diagnostic criteria that attribute symptoms to a pathogen which is detectable, whether it is or is not the aetiological agent of disease, may lead to outcome misclassification in epidemiological studies. Case–control studies are commonly undertaken to estimate vaccine effectiveness (VE) and present an opportunity to compare pathogen detection among individuals with and without clinically relevant symptoms. Considering this study context, we present a mathematical framework yielding simple estimators for the direct effects of vaccination on various aspects of host susceptibility. These include protection against acquisition of the pathogen of interest and protection against progression of this pathogen to disease following acquisition. We assess the impact of test sensitivity on these estimators and extend our framework to identify a ‘vaccine probe’ estimator for pathogen-specific aetiological fractions. We also derive biases affecting VE estimates under the test-negative design, a special case enrolling only symptomatic persons. Our results provide strategies for estimating pathogen-specific VE in the absence of a diagnostic gold standard. These approaches can inform the design and analysis of studies addressing numerous pathogens and vaccines.

## 1. Background

Attribution of infectious disease syndromes to a specific microbiological agent often involves detection of this agent, or its genetic material, among individuals experiencing relevant clinical signs and/or symptoms. However, in certain instances, detection of a pathogenic organism from a symptomatic individual does not provide a specific determination of disease aetiology. Such is the case for pathogens that may be shed or carried in a subclinical state as part of their natural history. For instance, influenza viruses, respiratory syncytial virus, human rhinoviruses and metapneumoviruses, and bacteria such as *Streptococcus pneumoniae*, *Haemophilus influenzae* and *Staphylococcus aureus* may each be found in respiratory tract specimens from individuals experiencing acute symptoms as well as individuals without clinically apparent illness, especially among children [1,2]. Similar circumstances arise with enteric viruses such as rotavirus, norovirus and sapovirus and bacteria such as *Shigella* and *Escherichia coli*, which may cause diarrhoea or colonize the gut asymptotically [3]. Likewise, depending on world region, malaria parasites, arboviruses, *Rickettsia* and various Gram-negative bacteria may be detectable among individuals experiencing acute febrile illness as well as those without symptoms [4]. Because many or most acquisitions of these pathogens do not progress to disease, only a proportion of symptomatic individuals among whom these pathogens are detectable may in fact be experiencing pathogen-attributable illness [5].

This circumstance creates challenges for studies of the effectiveness of vaccines and other pathogen-specific interventions. Whereas such studies typically aim to assess whether an intervention prevents a well-defined endpoint of disease attributable to the pathogen of interest, the absence of a gold-standard diagnostic tool may lead to misclassification of individuals' outcomes with respect to disease aetiology. Recently, large-scale studies of pneumonia [6] and diarrhoea [7] aetiology among children have employed case-control frameworks, enrolling symptomatic cases and community controls to estimate pathogen-attributable fractions of disease burden. The inferential methods of these studies centre on a comparison of the prevalence of each pathogen among symptomatic cases and asymptomatic controls [8–10]. While similar case-control frameworks may present an opportunity to correct for misclassification when estimating vaccine effectiveness (VE; [11–15]), such approaches have not been routinely incorporated into real-world studies.

Here we introduce strategies for estimation of VE against pathogen-specific endpoints in case-control studies. As estimands, we consider vaccine direct effect parameters that compare counterfactual outcomes for an individual who does or does not receive vaccination, as introduced by Halloran and colleagues in previous work [16]. These effects address the biological reduction in susceptibility conferred by an individual's response to vaccination, including protection against shedding or carriage of the pathogen, protection against progression of the pathogen to disease (conditioned on acquisition) and the cumulative result of these two forms of protection.

## 2. Study design

### 2.1. Framework

Consider a case-control study that enrolls cases experiencing a clinical syndrome of interest, which may be preventable by vaccination against one of several causative pathogens. Consider that the study enrolls controls based on clinical criteria unrelated to the pathogen of interest and to vaccination. Such controls may be asymptomatic persons or those experiencing an 'alternative' disease, which does not affect and is not affected by acquisition of the pathogen of interest or vaccination [17]. Assume that all or a proportion of individuals enrolled (including both cases and controls) are tested to determine the presence or the absence of the pathogen of interest.

For the methods described below, we envision a study where there is no *a priori* basis for determining whether a pathogen, once detected, is the true cause of symptoms for an individual patient. We also present simplifications of the approach for contexts where pathogen detection provides a definitive determination of disease aetiology.

### 2.2. Examples

Case-control studies involving the detection of enteropathogens from diarrhoea cases and asymptomatic controls [7], or the detection of respiratory viruses and bacteria in the upper respiratory tract of cases with and controls without acute respiratory symptoms [6], present an ideal environment for application of the approaches detailed below. Studies of diarrhoea-causing pathogens are a compelling example because the gastrointestinal tract is the site of both colonization and disease for such organisms. Testing may be performed on

**Table 1.** Variables or observations.

variable	definition
$Y_i$	indicator of case ( $Y_i = 1$ ) or control ( $Y_i = 0$ ) status, as defined by clinical symptoms, for individual $i$
$X_i$	indicator of pathogen presence ( $X_i = 1$ ) or absence ( $X_i = 0$ ), for individual $i$
$Z_i$	indicator of vaccine receipt ( $Z_i = 1$ ) or non-receipt ( $Z_i = 0$ ), for individual $i$
$D_i$	indicator that a pathogen is detected ( $D_i = 1$ ) or not detected ( $D_i = 0$ ), for individual $i$

stool specimens regardless of the presence of symptoms, and the detection of a pathogen in diarrhoeal stool is not always a clear indication of its aetiological significance [9,18,19]. Similar concerns arise with detection of various pathogens in the bloodstream, urine or respiratory tract among individuals experiencing acute febrile illness in endemic settings [4]. Although these challenges may be mitigated for certain pathogens by the use of quantitative molecular diagnostic tests [5], such testing remains relatively uncommon in clinical practice.

In studies of paediatric pneumonia, inaccessibility of the site of infection (i.e. the lung) presents a unique challenge where the approaches we consider below may offer particular value. The upper respiratory tract (including the oropharynx and nasopharynx) is a readily accessible site for swabbing, but is also the site of colonization or shedding for many pathogens; thus, identification of a pathogen in the upper respiratory tract of a symptomatic individual is not an indication of its aetiological significance. While lower respiratory tract secretions may be desirable because they originate from the site of infection, contamination with organisms from the upper respiratory tract may undermine the benefit of obtaining such specimens [20].

## 3. Notation and theoretical framework

Here we lay out a formal framework relating observable variables to epidemiological parameters. Define  $Y_i$  as an indicator of the case ( $Y_i = 1$ ) or control ( $Y_i = 0$ ) status of individual  $i$  with respect to the clinical syndrome of interest. With  $Y_i = 1$ , the individual experiences disease, while with  $Y_i = 0$ , the individual is either asymptomatic or experiencing a control condition unrelated to the pathogen or vaccine being studied. Define  $X_i$  as an indicator of the presence ( $X_i = 1$ ) or the absence ( $X_i = 0$ ) of the pathogen of interest, and define  $Z_i$  as an indicator of the individual's vaccination status as vaccinated (exposed,  $Z_i = 1$ ) or unvaccinated (unexposed,  $Z_i = 0$ ; table 1).

We next consider epidemiological parameters describing the prevalence and natural history of the vaccine-preventable pathogen of interest among cases and controls (table 2). Define the prevalence of the pathogen in the absence of disease and vaccine-derived immunity as  $\pi = \Pr(X_i = 1 | Y_i = 0, Z_i = 0)$ ; for simplicity and consistency with prior literature [16], we refer to this state as 'infection', recognizing that the terms 'carriage' or 'colonization' may be preferred when discussing asymptomatic natural history states of certain (e.g. bacterial) pathogens.

**Table 2.** Parameters of the model.

parameter	definition	range
$\pi$	prevalence of the pathogen of interest among individuals not experiencing disease symptoms potentially related to this pathogen (equal to the expected prevalence among controls)	0–1
$\theta_S$	prevalence ratio of the pathogen of interest among vaccinated versus unvaccinated control individuals not experiencing disease symptoms potentially related to this pathogen	0–1, for a protective vaccine
$\rho$	per-acquisition risk for progression of the pathogen of interest to cause disease among unvaccinated individuals	0–1
$\theta_P$	risk ratio for progression of the pathogen of interest to disease, given acquisition among vaccinated versus unvaccinated individuals	0–1, for a protective vaccine
$\omega$	risk (measured over the same period as $\rho$ ) of disease due to all causes other than the pathogen of interest	0–1
$\nu$	proportion of individuals in the population who are vaccinated	0–1
$\lambda$	rate of acquisition of the pathogen of interest among unvaccinated individuals not currently shedding or carrying this pathogen	greater than 0
$\gamma$	rate of clearance of the pathogen of interest among unvaccinated individuals, following acquisition	greater than 0
$a$	hazard ratio of acquisition of the pathogen of interest among vaccinated versus unvaccinated individuals not currently shedding or carrying this pathogen	0–1, for a protective vaccine
$b$	hazard ratio of clearance of the pathogen of interest among vaccinated versus unvaccinated individuals, following acquisition	greater than 1, for a protective vaccine
$r_P$	rate of progression of the pathogen of interest to disease following acquisition by unvaccinated individuals	greater than 0
$r_\omega$	rate of onset of disease due to all causes other than the pathogen of interest	greater than 0
$c$	hazard ratio of progression of the pathogen of interest to disease among vaccinated versus unvaccinated individuals, following acquisition	0–1, for a protective vaccine
$\alpha_0$	test sensitivity for detection of the pathogen in an individual experiencing shedding or carriage, without symptoms attributable to the pathogen	0–1
$\alpha_1$	test sensitivity for detection of the pathogen in an individual experiencing symptoms attributable to the pathogen	0–1

Define  $\theta_S\pi$  as the prevalence of infection with the pathogen of interest among vaccinated persons not experiencing disease, with  $VE_S = 1 - \theta_S$  indicating the reduction in prevalence attributable to vaccine effects on pathogen acquisition or clearance. Here we consider prevalence comparisons rather than time-to-event comparisons owing to the nature of data available in case-control studies, which would not include longitudinal observations of acquisition and clearance of the pathogen. We address the interpretation of  $\theta_S$  with respect to vaccine effects on the acquisition and duration of infection in a later section.

Define  $\rho$  as the proportion of acquisitions resulting in symptomatic disease in the absence of vaccine-conferred protection or as the proportion of infected individuals experiencing pathogen-attributable symptoms at any time. Define  $\theta_P$  as the relative risk of progression by the pathogen of interest, given vaccination, so that the vaccine direct effect against progression is  $VE_P = 1 - \theta_P$ . Last, allow  $\omega$  to represent the risk (prevalence) of disease attributable to all other causes. We assume here that onset of disease due to causes other than the pathogen of interest is unrelated to either prior vaccination or infection involving the pathogen of interest. This

assumption is consistent with the fundamental framework of case-control designs, where a control condition is chosen which is unaffected by the exposure(s) of interest (e.g. vaccination or infection) in order to provide a ‘null’ reference against which the prevalence of this exposure among cases may be compared [21]. Accounting for disease attributable to the pathogen of interest and to all other pathogens, we have

$$\left. \begin{aligned} \Pr(Y_i = 1|X_i = 1, Z_i = 0) &= \omega + \rho, \\ \Pr(Y_i = 1|X_i = 1, Z_i = 1) &= \omega + \theta_P\rho, \\ \Pr(Y_i = 1|X_i = 0, Z_i = 0) &= \Pr(Y_i = 1|X_i = 0, Z_i = 1) = \omega. \end{aligned} \right\} \quad (3.1a-c)$$

Note that, in a case-control study, enrolment of individuals with known outcome status  $Y_i = 1$  or  $Y_i = 0$  hinders direct measurement of  $\Pr(Y_i = y|X_i = x, Z_i = z)$ . Thus, we consider estimation frameworks below that do not require prospective observation of the probabilities presented in equation (3.1). As our study considers acute diseases (e.g. pneumonia, severe diarrhoea) that have low point prevalence within whole populations [22],  $\omega + \rho \approx 1 - (1 - \omega)(1 - \rho)$ .

We derive  $\rho$  and  $\omega$  with respect to rate parameters in a later section to relate  $\theta_p$  to the hazard ratio of progression of the pathogen of interest to symptomatic disease, given vaccination, in a competing risks framework.

#### 4. Vaccine direct effect against susceptibility ( $VE_S$ )

If an individual's control status is unrelated to risk of shedding the pathogen of interest, the relative risk of detection of the pathogen of interest among controls, given vaccination, provides the most straightforward basis for estimating  $VE_S$ ,

$$RR_S = \frac{\Pr(X_i = 1|Y_i = 0, Z_i = 1)}{\Pr(X_i = 1|Y_i = 0, Z_i = 0)} = \frac{\theta_S \pi}{\pi} = \theta_S = 1 - VE_S. \quad (4.1)$$

As an alternative,  $VE_S$  can be estimated by comparing the odds of vaccination and pathogen detection among controls against the 'null' odds of vaccination within the control group, irrespective of pathogen detection [23]. Such an approach may be advantageous under certain design considerations, most notably when pathogen detection data are available only from a subset of all control individuals from whom vaccination data are available (e.g. in the case of expensive or invasive diagnostic procedures [24,25]). Here, the odds ratio (OR) can be expressed as

$$OR_S = \frac{[\Pr(Z_i = 1, X_i = 1|Y_i = 0)][\Pr(Z_i = 0|Y_i = 0)]}{[\Pr(Z_i = 0, X_i = 1|Y_i = 0)][\Pr(Z_i = 1|Y_i = 0)]}. \quad (4.2)$$

To formalize this intuition, define  $v$  as the proportion of individuals receiving vaccination in the population; provided vaccination is independent of control status, we have  $\Pr(Z_i = 1|Y_i = 0) = v$ . From the assumption that control status is independent of vaccination and shedding of the pathogen of interest

$$\Pr(Z_i = 1, X_i = 1|Y_i = 0) = \Pr(X_i = 1|Z_i = 1, Y_i = 0) \Pr(Z_i = 1|Y_i = 0) = \theta_S \pi v \quad (4.3a)$$

and

$$\Pr(Z_i = 0, X_i = 1|Y_i = 0) = \Pr(X_i = 1|Z_i = 0, Y_i = 0) \Pr(Z_i = 0|Y_i = 0) = \pi(1 - v). \quad (4.3b)$$

Substituting into equation (4.2),

$$OR_S = \frac{[\theta_S \pi v] [(1 - v)]}{[\pi(1 - v)][v]} = \theta_S = 1 - VE_S, \quad (4.4)$$

recovering the same effect we would estimate using the risk ratio  $RR_S$ .

#### 5. Vaccine direct effect against progression ( $VE_P$ )

Here we address the vaccine-attributable reduction in susceptibility to disease, given a pathogen has overcome a host's vaccine-derived protection against acquisition of infection. To estimate this effect, consider first the OR of detection of the pathogen of interest, given symptoms ( $OR_p$ ). We expect this value to exceed 1 if acquisition of the pathogen increases individuals' risk of experiencing symptoms (i.e.  $\rho > 0$ ; [8,9]).

Among the unvaccinated, the relative odds of pathogen detection given the presence of symptoms is

$$OR_p^0 = \frac{\Pr(X_i = 1|Y_i = 1, Z_i = 0) \Pr(X_i = 0|Y_i = 0, Z_i = 0)}{\Pr(X_i = 0|Y_i = 1, Z_i = 0) \Pr(X_i = 1|Y_i = 0, Z_i = 0)}, \quad (5.1a)$$

while among the vaccinated,

$$OR_p^1 = \frac{\Pr(X_i = 1|Y_i = 1, Z_i = 1) \Pr(X_i = 0|Y_i = 0, Z_i = 1)}{\Pr(X_i = 0|Y_i = 1, Z_i = 1) \Pr(X_i = 1|Y_i = 0, Z_i = 1)}. \quad (5.1b)$$

We may expect that  $OR_p^1 < OR_p^0$  if vaccination confers protection against progression of the pathogen from infection to disease, thereby attenuating the association between pathogen detection and symptoms among the vaccinated. Of the terms appearing in equation (5.1), we have defined  $\Pr(X_i = 1|Y_i = 0, Z_i = 1)$  and  $\Pr(X_i = 1|Y_i = 0, Z_i = 0)$  above in equation (4.1). From the progression parameters  $\rho$ ,  $\omega$  and  $\theta_p$  we may further define the probability of pathogen detection, among cases, as

$$\Pr(X_i = 1|Y_i = 1, Z_i = 0) = \frac{\pi(\rho + \omega)}{\pi\rho + \omega} \quad (5.2a)$$

and

$$\Pr(X_i = 1|Y_i = 1, Z_i = 1) = \frac{\theta_S \pi(\theta_p \rho + \omega)}{(\theta_S \pi \theta_p \rho + \omega)}. \quad (5.2b)$$

As the complements of equations (5.2a) and (5.2b), respectively,

$$\Pr(X_i = 0|Y_i = 1, Z_i = 0) = \frac{(1 - \pi)\omega}{\pi\rho + \omega} \quad (5.2c)$$

and

$$\Pr(X_i = 0|Y_i = 1, Z_i = 1) = \frac{(1 - \theta_S \pi)\omega}{\theta_S \pi \theta_p \rho + \omega}. \quad (5.2d)$$

Substituting terms from equations (4.1) and (5.2) into equation (5.1), we have

$$OR_p^0 = \frac{[\pi(\rho + \omega)][(1 - \pi)]}{[\omega(1 - \pi)][\pi]} = \frac{\rho + \omega}{\omega} \quad (5.3a)$$

and

$$OR_p^1 = \frac{[\theta_S \pi(\theta_p \rho + \omega)][(1 - \theta_S \pi)]}{[(1 - \theta_S \pi)\omega][\theta_S \pi]} = \frac{\theta_p \rho + \omega}{\omega}. \quad (5.3b)$$

Here,  $OR_p^1 < OR_p^0$  for  $\theta_p < 1$  and  $\rho > 0$ ,  $\omega > 0$ . Rearranging the terms from equation (5.3) reveals  $\rho = \omega(OR_p^0 - 1)$  and  $\theta_p \rho = \omega(OR_p^1 - 1)$ . Substituting for  $\theta_p$

$$VE_P = 1 - \theta_p = 1 - \frac{OR_p^1 - 1}{OR_p^0 - 1}. \quad (5.4)$$

Thus, we may estimate the vaccine direct effect against progression of the targeted pathogen to disease by comparing the association of detection of the pathogen with the presence of symptoms, among vaccinated and unvaccinated individuals.



## 6. Vaccine direct effect against disease ( $VE_D$ , including acquisition and progression)

We last consider the cumulative extent of vaccine-conferred protection resulting from prevention of infection, and prevention of progression to disease in the event of breakthrough acquisition ( $VE_D$ ). From the joint probabilities of pathogen acquisition and progression among the vaccinated and unvaccinated,

$$\begin{aligned} VE_D &= 1 \\ &= \frac{\Pr(\text{pathogen-attributable disease}|\text{vaccinated})}{\Pr(\text{pathogen-attributable disease}|\text{unvaccinated})} \\ &= 1 - \frac{\theta_S \pi \theta_P \rho}{\pi \rho} = 1 - \theta_S \theta_P, \end{aligned} \quad (6.1)$$

which may thus be estimated as

$$VE_D = 1 - RR_S \times \frac{OR_p^1 - 1}{OR_p^0 - 1} \quad (6.2a)$$

or

$$VE_D = 1 - OR_S \times \frac{OR_p^1 - 1}{OR_p^0 - 1}, \quad (6.2b)$$

as defined above.

## 7. Interpretation of risks and rates

### 7.1. Pathogen shedding or carriage parameters

Interpretation of pathogen prevalence measures presents a challenge in studies of VE against endpoints such as asymptomatic infection due to the dynamic nature of transmission [16,26,27]. Here we relate  $\theta_S = 1 - VE_S$  to vaccine direct effects on the rates of pathogen acquisition and clearance.

Define  $\lambda$  and  $\gamma$  as the rates at which unvaccinated individuals acquire and clear the pathogen of interest, respectively. Taking  $\pi$  to indicate the equilibrium prevalence of shedding or colonization, with  $\pi > 0$ , we have  $\lambda(1 - \pi) = \gamma\pi$  and  $\pi = \lambda / (\lambda + \gamma)$ . Further defining  $a$  and  $b$  as the hazard ratios of acquisition and clearance of the pathogen among vaccinated persons relative to unvaccinated persons (table 2), we have  $a\lambda(1 - \theta_S\pi) = b\gamma\theta_S\pi$ . Substituting for  $\pi$ , we obtain

$$\theta_S = \frac{a(\lambda + \gamma)}{a\lambda + b\gamma}. \quad (7.1a)$$

If the rate of clearance of the pathogen far exceeds the rate of acquisition ( $\gamma \gg \lambda$ ), we have

$$\theta_S \cong \lim_{\lambda/\gamma \rightarrow 0} (a(\lambda + \gamma)/(a\lambda + b\gamma)) = a/b. \quad (7.1b)$$

Thus, the prevalence ratio  $\theta_S$  among controls (as estimated by  $RR_S$  or  $OR_S$ ) may be interpretable as the hazard ratio of acquisition if  $b=1$  (such that vaccination does not affect time to clearance of the pathogen) or as the inverse of the hazard ratio of clearance if  $a=1$  (such that vaccination does not affect acquisition of the pathogen).

### 7.2. Disease progression parameters

We may relate the risk parameters  $\rho$ ,  $\omega$  and  $\theta_P$  to epidemiological rates under a similar intuition. Define  $r_p$  as the rate of progression of the pathogen of interest to disease, following acquisition by an unvaccinated individual, and define  $r_\omega$  as the rate of onset of

disease due to all other causes. Here we may use a competing risks framework to define  $\rho$  and  $\omega$  as the probability that disease due to each cause precedes disease due to the other cause, or clearance of the pathogen. Defining  $f(t|r)$  as the density function for an event with rate  $r$  occurring at time  $t$ , we have

$$\rho = \int_0^\infty f(t|r_p)(1 - F(t|r_\omega + \gamma)) dt \quad (7.2a)$$

and

$$\omega = \int_0^\infty f(t|r_\omega)(1 - F(t|r_p + \gamma)) dt. \quad (7.2b)$$

Next defining  $c$  as the hazard ratio of progression of the pathogen of interest to disease following acquisition (comparing vaccinated with unvaccinated individuals), we have

$$\theta_P \rho = \int_0^\infty f(t|cr_p)(1 - F(t|r_\omega + \gamma)) dt. \quad (7.2c)$$

Assuming exponentially distributed onset times, we have  $\rho = r_p/(r_p + r_\omega + \gamma)$ ,  $\omega = r_\omega/(r_p + r_\omega + \gamma)$  and  $\theta_P \rho = cr_p/(cr_p + r_\omega + \gamma)$ , so that

$$\theta_P = \frac{c(r_p + r_\omega + \gamma)}{cr_p + r_\omega + \gamma}. \quad (7.3)$$

If the rate of progression of the pathogen of interest is well below the rate of disease onset due to all other causes ( $r_\omega \gg r_p$ ), or the rate of clearance of the pathogen ( $\gamma \gg r_p$ ), we have  $\theta_P \cong \lim_{r_p/(r_\omega + \gamma) \rightarrow 0} (c(r_p + r_\omega + \gamma)/(cr_p + r_\omega + \gamma)) = c$ .

Thus, for diseases that are uncommon,  $OR_P = \theta_P$  may be interpretable as the hazard ratio of disease progression, following acquisition, for a vaccinated versus unvaccinated individual.

## 8. Studies with high diagnostic specificity in disease

The framework we introduce above also has value for estimating vaccine-conferred protection against progression to endpoints where the distinction between true disease and subclinical shedding or carriage is unambiguous. For instance, detection of commensal bacteria in ordinarily sterile body fluids such as the bloodstream or cerebrospinal fluid is a defining characteristic of invasive infections involving these organisms. Detection of respiratory viruses in the bloodstream may similarly provide a marker of very severe disease attributable to these pathogens [28–31]. Even when the isolate source is not an indicator of disease aetiology, pathogen abundance may inform the distinction between aetiological and non-aetiological detections [5].

In the event that we may exclude alternative aetiologies among cases in whom the pathogen is detected, the probability of pathogen detection, among cases (equations (5.2a, b)), may be rewritten as

$$\Pr(X_i = 1|Y_i = 1, Z_i = 0) = \frac{\pi\rho}{\pi\rho + (1 - \pi)\omega} \quad (8.1a)$$

and

$$\Pr(X_i = 1|Y_i = 1, Z_i = 1) = \frac{\theta_S \pi \theta_P \rho}{\theta_S \pi \theta_P \rho + (1 - \theta_S \pi)\omega}, \quad (8.1b)$$

respectively. Here, we subtract  $\pi\omega(1 - v)$  and  $\theta_S\pi\omega v$  from our earlier definitions of  $\Pr(Y_i = 1, Z_i = 0)$  and  $\Pr(Y_i = 1, Z_i = 1)$ ,

respectively, because episodes of disease not attributable to the pathogen of interest among cases shedding or carrying this pathogen are not observed. Thus,

$$\text{OR}_p^0 = \frac{[\pi\rho][(1-\pi)]}{[(1-\pi)\omega][\pi]} = \frac{\rho}{\omega} \quad (8.2a)$$

and

$$\text{OR}_p^1 = \frac{[\theta_S\pi\theta_P\rho][(1-\theta_S\pi)]}{[(1-\theta_S\pi)\omega][\theta_S\pi]} = \frac{\theta_P\rho}{\omega}, \quad (8.2b)$$

so that

$$\text{VE}_P = 1 - \theta_P = 1 - \frac{\text{OR}_p^1}{\text{OR}_p^0}. \quad (8.3)$$

We note from equations (5.4) and (8.3) that estimates of  $\text{VE}_P$  may be most stable when  $\text{OR}_p^0 \gg 1$  (which may also allow  $\text{OR}_p^1 \gg 1$  if  $\theta_P$  is not close to zero). In this regard, the analysis approaches we consider here may be most reliable for pathogens that play an important aetiological role in the disease syndromes under investigation, resulting in high ORs.

The special case of high diagnostic specificity in disease further allows  $\text{VE}_D$  to be estimated directly by comparing the odds of prior vaccination among symptomatic cases with the pathogen detected versus asymptomatic controls among whom the pathogen is not detected. We show  $\Pr(Z_i = z|X_i = 1, Y_i = 0)$  in equation (4.3). Similarly,

$$\begin{aligned} \Pr(Z_i = 1|X_i = 1, Y_i = 1) &= \frac{\Pr(X_i = 1, Y_i = 1|Z_i = 1)\Pr(Z_i = 1)}{\Pr(X_i = 1, Y_i = 1)} \\ &= \frac{\theta_S\theta_P v}{\theta_S\theta_P v + (1-v)} \end{aligned} \quad (8.4a)$$

and its complement

$$\Pr(Z_i = 0|X_i = 1, Y_i = 1) = \frac{1-v}{\theta_S\theta_P v + (1-v)}, \quad (8.4b)$$

under the circumstance that the pathogen, if detected in a symptomatic individual, is the true cause of disease. Taken together, the vaccine effect can be expressed as

$$\begin{aligned} \text{VE}_D &= 1 - \frac{\Pr(Z_i = 1|X_i = 1, Y_i = 1)\Pr(Z_i = 0|X_i = 0, Y_i = 0)}{\Pr(Z_i = 0|X_i = 1, Y_i = 1)\Pr(Z_i = 1|X_i = 0, Y_i = 0)} \\ &= 1 - \frac{[\theta_S\theta_P v][(1-\pi)(1-v)]}{[(1-v)][(1-\pi)v]} \\ &= 1 - \theta_S\theta_P. \end{aligned} \quad (8.5)$$

## 9. Bias under imperfect test sensitivity

For certain conditions, the sensitivity of pathogen detection approaches may differ for individuals experiencing pathogen-attributable disease versus subclinical shedding or colonization. For instance, higher pathogen load among individuals experiencing disease may improve the probability of pathogen detection among true disease cases via certain diagnostic assays [5,32]. Alternatively, for conditions such as pneumonia involving infection at difficult-to-sample anatomical sites, shedding or carriage at the sampled site (e.g. oronasopharynx) may cease before the clearance of the pathogen from the site of infection [33–35].

Take  $D_i = 1$  and  $D_i = 0$  to indicate detection or non-detection of the pathogen of interest (table 1). To account

for test sensitivity in the formulations introduced above, we may replace  $X_i$  with  $D_i$  in our previous equations, so that

$$\text{RR}_S = \frac{\Pr(D_i = 1|Y_i = 0, Z_i = 1)}{\Pr(D_i = 1|Y_i = 0, Z_i = 0)} \quad (9.1a)$$

and

$$\text{OR}_S = \frac{[\Pr(Z_i = 1, D_i = 1|Y_i = 0)][\Pr(Z_i = 0|Y_i = 0)]}{[\Pr(Z_i = 0, D_i = 1|Y_i = 0)][\Pr(Z_i = 1|Y_i = 0)]} \quad (9.1b)$$

for estimation of  $\text{VE}_S$ , while

$$\text{OR}_P^0 = \frac{\Pr(D_i = 1|Y_i = 1, Z_i = 0)\Pr(D_i = 0|Y_i = 0, Z_i = 0)}{\Pr(D_i = 0|Y_i = 1, Z_i = 0)\Pr(D_i = 1|Y_i = 0, Z_i = 0)} \quad (9.2a)$$

and

$$\text{OR}_P^1 = \frac{\Pr(D_i = 1|Y_i = 1, Z_i = 1)\Pr(D_i = 0|Y_i = 0, Z_i = 1)}{\Pr(D_i = 0|Y_i = 1, Z_i = 1)\Pr(D_i = 1|Y_i = 0, Z_i = 1)} \quad (9.2b)$$

for estimation of  $\text{VE}_D$ .

Define  $\alpha_1$  and  $\alpha_0$  as the sensitivity of the detection method—i.e. the probability of detecting the pathogen, given it is present—for individuals who are ( $\alpha_1$ ) and are not ( $\alpha_0$ ) experiencing disease attributable to this pathogen. The expected prevalence of pathogen detection among those without symptoms is thus  $\Pr(D_i = 1|Y_i = 0) = \alpha_0 \Pr(X_i = 1|Y_i = 0)$ . Substituting into equation (5.2) reveals no effect on  $\text{RR}_S$  and  $\text{OR}_S$ , as  $\alpha_0$  cancels in the numerator and denominator.

Next considering  $\text{OR}_P^0$ , the detection probability among unvaccinated cases is

$$\begin{aligned} \Pr(D_i = 1|Y_i = 1, Z_i = 0) &= \frac{\Pr(Y_i = 1, Z_i = 0|X_i = 1, D_i = 1)\Pr(X_i = 1, D_i = 1)}{\Pr(Y_i = 1, Z_i = 0)} \end{aligned} \quad (9.3a)$$

since  $\Pr(X_i = 0, D_i = 1) = 0$ . Among individuals infected with the pathogen of interest, detection and disease may jointly occur with probability  $\alpha_1\rho$  for disease due to the pathogen of interest, and with probability  $\alpha_0\omega$  for disease due to other causes, so that

$$\begin{aligned} \Pr(Y_i = 1, Z_i = 0|X_i = 1, D_i = 1)\Pr(X_i = 1, D_i = 1) &= \pi(\alpha_1\rho + \alpha_0\omega) \end{aligned} \quad (9.3b)$$

and

$$\Pr(D_i = 1|Y_i = 1, Z_i = 0) = \frac{\pi(\alpha_1\rho + \alpha_0\omega)}{\pi\rho + \omega}. \quad (9.3c)$$

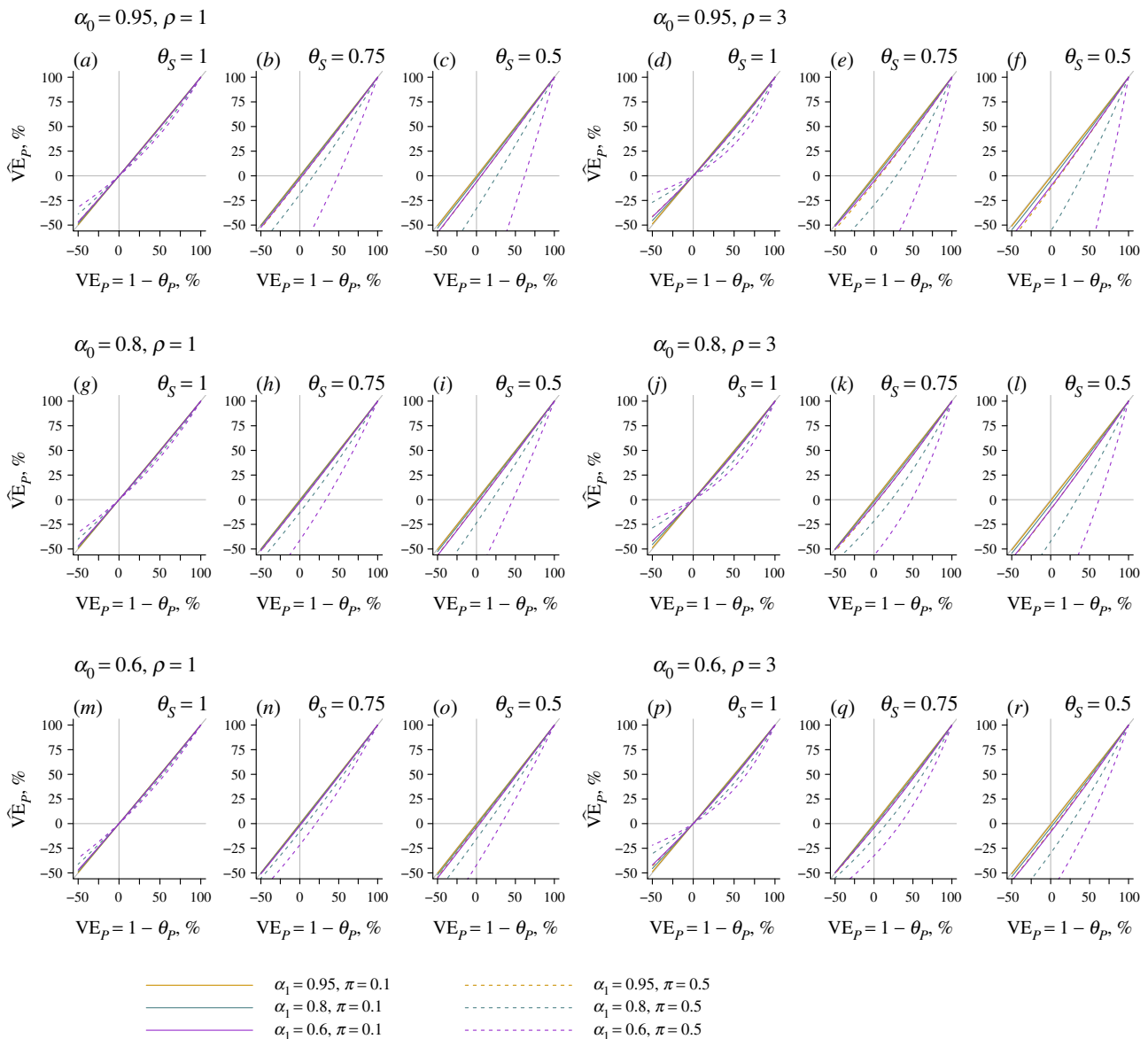
The probability of no pathogen detection among those with disease is

$$\Pr(D_i = 0|Y_i = 1, Z_i = 0) = \frac{\pi\rho(1-\alpha_1) + \omega(1-\alpha_0\pi)}{\pi\rho + \omega}, \quad (9.3d)$$

which we may obtain from the joint probabilities of disease due to the pathogen of interest and detection failure,  $\pi\rho(1-\alpha_1)$ , and disease due to all other causes without pathogen detection,  $\omega(1-\alpha_0\pi)$ .

Similarly, among the vaccinated,

$$\Pr(D_i = 1|Y_i = 1, Z_i = 1) = \frac{\theta_S\pi(\alpha_1\theta_P\rho + \alpha_0\omega)}{\theta_S\pi\theta_P\rho + \omega} \quad (9.4a)$$



**Figure 1.** Effect of imperfect test sensitivity on estimates of vaccine-conferred protection against progression of the pathogen to disease. Using the expressions for  $OR_p^0$  and  $OR_p^1$  in equation (9.5), we plot the estimated effect of vaccination on risk of progression of the pathogen of interest, given acquisition ( $\widehat{VE}_p = 1 - (OR_p^1 - 1)/(OR_p^0 - 1)$ ) under differing conditions of sensitivity of detection of the pathogen when it is or is not the agent of disease ( $\alpha_1$  and  $\alpha_0$  values of 0.95, 0.8 and 0.6). The  $x$ -axis indicates the true vaccine effect against progression; departures from the 1:1 diagonal (grey) line indicate bias. Solid and dashed lines correspond to estimates assuming 10% and 50% prevalence, respectively, of the pathogen of interest among unvaccinated individuals without symptoms. Within each row, the left three panels assume acquisition of the pathogen doubles individuals' risk of all-cause disease ( $\rho/\omega = 1$ ), while the right three panels assume acquisition of the pathogen results in a fourfold increase in individuals' risk of all-cause disease ( $\rho/\omega = 3$ ).

and

$$\Pr(D_i = 0 | Y_i = 1, Z_i = 1) = \frac{\theta_S \pi \theta_P \rho (1 - \alpha_1) + \omega (1 - \alpha_0 \theta_S \pi)}{\theta_S \pi \theta_P \rho + \omega}. \quad (9.4b)$$

Substituting into equation (8.2), we have

$$\begin{aligned} OR_p^0 &= \frac{[\pi(\alpha_1 \rho + \alpha_0 \omega)][1 - \alpha_0 \pi]}{[\pi \rho (1 - \alpha_1) + \omega (1 - \alpha_0 \pi)][\alpha_0 \pi]} \\ &= \frac{(\alpha_1 \rho + \alpha_0 \omega)(1 - \alpha_0 \pi)}{\alpha_0 [\pi \rho (1 - \alpha_1) + \omega (1 - \alpha_0 \pi)]} \end{aligned} \quad (9.5a)$$

and

$$\begin{aligned} OR_p^1 &= \frac{[\theta_S \pi (\alpha_1 \theta_P \rho + \alpha_0 \omega)][1 - \alpha_0 \theta_S \pi]}{[\theta_S \pi \theta_P \rho (1 - \alpha_1) + \omega (1 - \alpha_0 \theta_S \pi)][\alpha_0 \theta_S \pi]} \\ &= \frac{(\alpha_1 \theta_P \rho + \alpha_0 \omega)(1 - \alpha_0 \theta_S \pi)}{\alpha_0 [\theta_S \pi \theta_P \rho (1 - \alpha_1) + \omega (1 - \alpha_0 \theta_S \pi)]}. \end{aligned} \quad (9.5b)$$

Here,  $(OR_p^1 - 1)/(OR_p^0 - 1) = \theta_P$  if  $\alpha_1 = 1$ ; otherwise,  $(OR_p^1 - 1)/(OR_p^0 - 1) > \theta_P$  for  $0 < \theta_P < 1$ , over the ranges of all other parameters listed in table 2. Thus, if vaccination confers protection against disease progression, imperfect sensitivity affecting pathogen detection among symptomatic cases is expected to lead to underestimation of the true effect. We illustrate the magnitude of this bias in figure 1 under various parametrizations. Bias increases with higher prevalence of the pathogen among persons without symptoms ( $\pi$ ), with higher rates of progression of the pathogen to disease ( $\rho/\omega$ ) and with lower sensitivity for pathogen detection in disease ( $\alpha_1$ ). Interestingly, bias also worsens as  $\alpha_0$  approaches 1.

The nature and extent of bias further depends upon whether vaccination confers protection against pathogen shedding or carriage ( $\theta_S$ ). With  $\theta_S = 1$  (i.e.  $VE_S = 0$ ), the estimate  $\widehat{VE}_p$



converges to the true value when  $VE_p = 0\%$  or  $VE_p = 100\%$ , and is sign-unbiased, such that  $\widehat{VE}_p > 0\%$  when  $VE_p > 0\%$ , and  $\widehat{VE}_p < 0\%$  when  $VE_p < 0\%$ . By contrast, the estimate may be sign-biased when vaccination confers protection against shedding (i.e.  $0 < \theta_s < 1$ ); here  $\widehat{VE}_p$  might suggest increased risk of disease progression among vaccinated individuals who acquire the pathogen, when in fact vaccination protects against disease progression. However, this circumstance is expected only under extreme scenarios, where shedding or carriage is very prevalent among asymptomatic individuals, true protection against disease progression is very weak and test sensitivity is very high in the context of subclinical shedding or carriage yet low among those experiencing disease. This latter circumstance is particularly unlikely to be met, as pathogen shedding (and thus likelihood of detection) is often associated with the presence of symptoms [5].

## 10. Bias under test-negative design approaches

### 10.1. Protection against disease

The test-negative design (TND) has come into routine use in studies of VE [36]. Studies using this design estimate vaccine direct effects by comparing the odds of prior vaccination among individuals experiencing clinically apparent illness who test positive or negative for a pathogen of interest ( $OR_D^{TND}$ ): in contrast to traditional case-control studies, all subjects experience the clinical syndrome of interest. Case-control studies such as we have considered above may resemble TND studies in that receipt of a test is independent of the likelihood of detecting the pathogen of interest [15,37,38]; however, 'control' subjects in TND studies are typically defined as individuals who experience disease without detection of the pathogen of interest. It has been recognized previously that detections of non-aetiological pathogens may influence estimates of VE under TND studies, although the quantitative extent of resulting bias is not well understood [11,13,15]. Here, we derive the mathematical form of this bias using the framework described above.

Studies using the TND aim to estimate the vaccine direct effect against disease attributable to the pathogen of interest according to

$$\begin{aligned} \widehat{VE}_D &= 1 - OR_D^{TND} \\ &= 1 - \frac{\Pr(Z_i = 1|X_i = 1, Y_i = 1) \Pr(Z_i = 0|X_i = 0, Y_i = 1)}{\Pr(Z_i = 0|X_i = 1, Y_i = 1) \Pr(Z_i = 1|X_i = 0, Y_i = 1)}. \end{aligned} \quad (10.1)$$

Here we may define

$$\begin{aligned} \Pr(Z_i = 1|X_i = 1, Y_i = 1) &= \frac{\Pr(X_i = 1, Y_i = 1|Z_i = 1) \Pr(Z_i = 1)}{\Pr(X_i = 1, Y_i = 1)} \\ &= \frac{\theta_s(\theta_p \rho + \omega)v}{\theta_s(\theta_p \rho + \omega)v + (\rho + \omega)(1 - v)} \end{aligned} \quad (10.2a)$$

and

$$\Pr(Z_i = 0|X_i = 1, Y_i = 1) = \frac{(\rho + \omega)(1 - v)}{\theta_s(\theta_p \rho + \omega)v + (\rho + \omega)(1 - v)}. \quad (10.2b)$$

Similarly, we have

$$\Pr(Z_i = 1|X_i = 0, Y_i = 1) = \frac{(1 - \theta_s \pi)v}{(1 - \theta_s \pi)v + (1 - \pi)(1 - v)} \quad (10.3a)$$

and

$$\Pr(Z_i = 0|X_i = 0, Y_i = 1) = \frac{(1 - \pi)(1 - v)}{(1 - \theta_s \pi)v + (1 - \pi)(1 - v)}. \quad (10.3b)$$

Substituting equations (10.2) and (10.3) into equation (10.1) to derive  $OR_D^{TND}$ ,

$$\begin{aligned} OR_D^{TND} &= \frac{[\theta_s(\theta_p \rho + \omega)v][(1 - \pi)(1 - v)]}{[(\rho + \omega)(1 - v)][(1 - \theta_s \pi)v]} \\ &= \frac{\theta_s(\theta_p \rho + \omega)(1 - \pi)}{(\rho + \omega)(1 - \theta_s \pi)}. \end{aligned} \quad (10.4)$$

This estimator approaches the true vaccine direct effect against acquisition and progression of the pathogen of interest ( $VE_D = 1 - \theta_s \theta_p$ ) under two conditions. First, if values of  $\pi$  and  $\omega$  (or the ratio  $\omega/\rho$ ) concurrently approach zero, and, second, if values of  $\theta_s$  and/or  $\theta_D$  approach zero (resulting in near-100% protection). We illustrate the quantitative extent of bias under differing conditions in figure 2.

### 10.2. Bias in the absence of diagnostic uncertainty

We also identify bias under the TND framework when detection of the pathogen from cases can be assumed to indicate true disease aetiology. Following equation (8.4), in this instance we may express the probability of vaccination among cases as

$$\begin{aligned} \Pr(Z_i = 1|X_i = 0, Y_i = 1) &= \frac{\Pr(X_i = 0, Y_i = 1|Z_i = 1) \Pr(Z_i = 1)}{\Pr(X_i = 0, Y_i = 1)} \\ &= \frac{(1 - \theta_s \pi)v}{(1 - \theta_s \pi)v + (1 - \pi)(1 - v)} \end{aligned} \quad (10.5a)$$

and

$$\Pr(Z_i = 0|X_i = 0, Y_i = 1) = \frac{(1 - \pi)(1 - v)}{(1 - \theta_s \pi)v + (1 - \pi)(1 - v)}. \quad (10.5b)$$

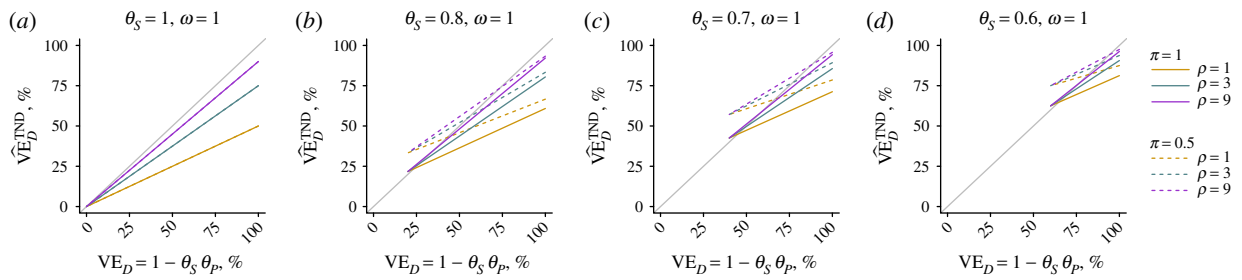
Using the expressions for  $\Pr(Z_i = 1|X_i = 1, Y_i = 1)$  and  $\Pr(Z_i = 0|X_i = 1, Y_i = 1)$  presented in equation (8.4), we have

$$OR_D^{TND} = \frac{[\theta_s \theta_p v][(1 - \pi)(1 - v)]}{[(1 - v)][(1 - \theta_s \pi)v]} = \frac{\theta_s \theta_p (1 - \pi)}{1 - \theta_s \pi}, \quad (10.6)$$

which converges to  $\theta_s \theta_p$  as  $\theta_s$ ,  $\theta_p$  or  $\pi$  approach zero. Otherwise, this approach will overestimate the magnitude of the true vaccine direct effect (i.e.  $OR_D^{TND} < \theta_s \theta_p$ ).

### 10.3. Protection against infection

Our mathematical framework reveals similar biases affecting estimation of  $VE_S$  under TND-like approaches that define cases as asymptomatic individuals among whom the pathogen is detected, and controls as asymptomatic individuals among whom the pathogen is not detected. Although TND studies typically enrol individuals experiencing a clinical syndrome that could be caused by the pathogen of interest, individuals may, in theory, be matched on any clinical status, including the absence of symptoms potentially attributable to the



**Figure 2.** Bias in estimated protection against disease under the TND. We plot the estimated effect of vaccination on risk of disease attributable to the pathogen of interest generated under the TND ( $\widehat{VE}_D^{\text{TND}}$ ), indicating the true vaccine direct effect ( $VE_D = 1 - \theta_S \theta_p$ ) on the  $x$ -axis. Departures from the 1 : 1 diagonal (grey) line indicate bias. Solid and dashed lines correspond to estimates assuming 10% and 50% prevalence, respectively, of the pathogen of interest among unvaccinated individuals without symptoms, with differing line colours corresponding to differing values of the relative risk of symptoms given acquisition of the pathogen ( $\rho/\omega$ ).

pathogen of interest. Here, investigators would aim to estimate VE against acquisitions that result only in subclinical infection by comparing the odds of prior vaccination among asymptomatic individuals who test positive (cases) or negative (controls) for the pathogen of interest [11,36,38]:

$$OR_S^{\text{TND}} = \frac{[\Pr(Z_i = 1, X_i = 1 | Y_i = 0)] [\Pr(Z_i = 0, X_i = 0 | Y_i = 0)]}{[\Pr(Z_i = 0, X_i = 1 | Y_i = 0)] [\Pr(Z_i = 1, X_i = 0 | Y_i = 0)]} \quad (10.7)$$

We have presented  $\Pr(Z_i = 1, X_i = 1 | Y_i = 0)$  and  $\Pr(Z_i = 0, X_i = 1 | Y_i = 0)$  in equation (4.3). We may similarly derive

$$\Pr(Z_i = 1, X_i = 0 | Y_i = 0) = \Pr(X_i = 0 | Z_i = 1, Y_i = 0) \Pr(Z_i = 1 | Y_i = 0) = (1 - \theta_S \pi) v \quad (10.8a)$$

and

$$\Pr(Z_i = 0, X_i = 0 | Y_i = 0) = \Pr(X_i = 0 | Z_i = 0, Y_i = 0) \Pr(Z_i = 0 | Y_i = 0) = (1 - \pi)(1 - v). \quad (10.8b)$$

Substituting into equation (10.7), we have

$$OR_S^{\text{TND}} = \frac{[\theta_S \pi v] [(1 - \pi)(1 - v)]}{[\pi(1 - v)] [(1 - \theta_S \pi) v]} = \frac{\theta_S(1 - \pi)}{1 - \theta_S \pi}, \quad (10.9)$$

revealing the same form of bias inherent to equation (10.6).

## 11. Attributable fraction estimation

Data from case–control studies investigating pathogen detection among symptomatic and asymptomatic persons may also be useful for assessing the fraction of cases attributable to the pathogen of interest. Studies addressing the effects of vaccination on non-specific endpoints provide a framework to ‘probe’ the proportion of cases attributable to the vaccine-preventable pathogen [39,40], based on effects against both all-cause and pathogen-specific disease outcomes.

The OR of vaccination among cases and controls, irrespective of pathogen detection, is

$$OR_A = \frac{\Pr(Z_i = 1 | Y_i = 1) \Pr(Z_i = 0 | Y_i = 0)}{\Pr(Z_i = 0 | Y_i = 1) \Pr(Z_i = 1 | Y_i = 0)}. \quad (11.1)$$

As in equation (4.2), we have  $\Pr(Z_i = 1 | Y_i = 0) = v$  and  $\Pr(Z_i = 0 | Y_i = 0) = 1 - v$ , while

$$\Pr(Z_i = 1 | Y_i = 1) = \frac{\Pr(Y_i = 1 | Z_i = 1) \Pr(Z_i = 1)}{\Pr(Y_i = 1)} = \frac{(\theta_S \pi \theta_p \rho + \omega) v}{(\theta_S \pi \theta_p \rho + \omega) v + (\pi \rho + \omega)(1 - v)} \quad (11.2a)$$

and

$$\Pr(Z_i = 0 | Y_i = 1) = \frac{(\pi \rho + \omega)(1 - v)}{(\theta_S \pi \theta_p \rho + \omega) v + (\pi \rho + \omega)(1 - v)}. \quad (11.2b)$$

Thus, we may express the effect of vaccination against all-cause disease as

$$OR_A = \frac{[(\theta_S \pi \theta_p \rho + \omega) v] [1 - v]}{[(\pi \rho + \omega)(1 - v)] [v]} = \frac{\theta_S \pi \theta_p \rho + \omega}{\pi \rho + \omega}. \quad (11.3)$$

We note that this result equals the risk ratio of disease given vaccination,  $\Pr(Y_i = 1 | Z_i = 1) / \Pr(Y_i = 1 | Z_i = 0)$ .

Rearranging the terms in equation (11.3), we obtain the risk (or rate) ratio of disease attributable to the pathogen of interest, versus other causes, as

$$\frac{\pi \rho}{\omega} = \frac{1 - OR_A}{OR_A - \theta_S \theta_p}, \quad (11.4a)$$

where  $1 - VE_D$  (equation (6.1)) may provide an unbiased input for  $\theta_S \theta_p$ . Thus, the proportion of disease attributable to the pathogen of interest, in the absence of vaccine-derived protection, is

$$\frac{\pi \rho}{\pi \rho + \omega} = \frac{1 - OR_A}{1 - \theta_S \theta_C}. \quad (11.4b)$$

The result is mathematically equivalent to previous formulations of vaccine probe estimators including estimated protection against non-specific and specific endpoints in the numerator and denominator, respectively [22,41].

## 12. Summary

Microbiological diagnosis of certain infectious diseases may be complicated by the prevalence of putative aetiological pathogens among individuals whose symptoms are due to other causes. Here we present a framework for estimating the direct effects of vaccination on different aspects of individuals’ susceptibility through the collection of pathogen detection data in case–control studies. A simple analytic extension provides a ‘vaccine probe’ framework for estimating the proportion of cases attributable to the pathogen of interest.

The approach we have taken also enables assessment of the quantitative extent of bias arising under different scenarios. We identify that pathogen-specific effects of vaccination may be underestimated when a pathogen is detected with low sensitivity among disease cases. Of the disease-specific applications we have described, this circumstance probably has the greatest importance for pneumonia, where pathogens

may continue causing disease in the lung after clearance from their commensal niche in the upper respiratory tract. This circumstance would presumably reduce the sensitivity of pathogen detection among disease cases to a greater extent than what is expected among controls.

We also identify that TND studies may produce biased VE estimates either with or without misclassification of symptomatic individuals based on pathogen detection. Notably, we identify that bias in TND studies may result in either over-estimation or under-estimation of true VE, making the interpretation of estimates difficult. Here, we do not consider additional issues that have been found to impact the reliability of estimates from TND studies, including individuals' acquisition of immunity following natural infection and the mode of vaccine action [38,42] (e.g. 'leaky' protection conferring partial risk reductions to all individuals, or 'all-or-none' protection among a proportion of individuals). Nonetheless, our findings add to a growing list of concerns about the uses of the TND for estimation of

VE [11,15,43]. The methods we lay out here provide an opportunity to correct for biases arising in conventional TND studies by enrolling control groups without symptoms potentially attributable to the pathogen of interest.

Case-control studies are commonly undertaken to estimate the direct effects of vaccination. Collecting pathogen detection data from controls, alongside symptomatic cases, presents a simple strategy to correct for misclassification of symptoms not attributable to a detected pathogen. Provided such data are collected, the approaches we describe permit estimation of VE against pathogen-specific endpoints without a diagnostic gold standard.

**Data accessibility.** Code to replicate the figures is available at <https://github.com/joelwnard/ccVE>.

**Competing interests.** I declare I have no competing interests.

**Funding.** JAL was supported by grant R01AI148127 from the National Institute of Allergy and Infectious Diseases (US National Institutes of Health).

## References

- Levine OS *et al.* 2012 The pneumonia aetiology research for child health project: a 21st century childhood pneumonia aetiology study. *Clin. Infect. Dis.* **54**, S93–S101. (doi:10.1093/cid/cir1052)
- Lewnard JA, Givon-Lavi N, Huppert A, Pettigrew MM, Regev-Yochay G, Dagan R, Weinberger DM. 2016 Epidemiological markers for interactions among *Streptococcus pneumoniae*, *Haemophilus influenzae*, and *Staphylococcus aureus* in upper respiratory tract carriage. *J. Infect. Dis.* **213**, 1596–1605. (doi:10.1093/infdis/jiv761)
- Kotloff KL *et al.* 2012 The Global Enteric Multicenter Study (GEMS) of diarrheal disease in infants and young children in developing countries: epidemiological and clinical methods of the case/control study. *Clin. Infect. Dis.* **55**, S232–S245. (doi:10.1093/cid/cis753)
- Tam PYI, Obaro SK, Storch G. 2016 Challenges in the aetiology and diagnosis of acute febrile illness in children in low- and middle-income countries. *J. Pediatric Infect Dis Soc.* **5**, 190–205. (doi:10.1093/jpids/piw016)
- Liu J *et al.* 2014 Development and assessment of molecular diagnostic tests for 15 enteropathogens causing childhood diarrhoea: a multicentre study. *Lancet Infect. Dis.* **14**, 716–724. (doi:10.1016/S1473-3099(14)70808-4)
- O'Brien KL *et al.* 2019 Causes of severe pneumonia requiring hospital admission in children without HIV infection from Africa and Asia: the PERCH multi-country case-control study. *Lancet* **394**, 757–779. (doi:10.1016/S0140-6736(19)30721-4)
- Kotloff KL *et al.* 2013 Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study. GEMS): a prospective, case-control study. *Lancet* **382**, 209–222. (doi:10.1016/S0140-6736(13)60844-2)
- Bruzzi P, Green SB, Byar DP, Brinton L, Schairer C. 1985 Estimating the population attributable risk for multiple risk factors using case-control data. *Am. J. Epidemiol.* **122**, 904–914. (doi:10.1093/oxfordjournals.aje.a114174)
- Blackwelder WC, Biswas K, Wu Y, Kotloff KL, Farag TH, Nasrin D, Graubard BI, Sommerfelt H, Levine MM. 2012 Statistical methods in the Global Enteric Multicenter Study (GEMS). *Clin. Infect. Dis.* **55**, S246–S253. (doi:10.1093/cid/cis788)
- Wu Z, Deloria-Knoll M, Zeger SL. 2017 Nested partially latent class models for dependent binary data; estimating disease aetiology. *Biostatistics* **18**, 200–213. (doi:10.1093/biostatistics/kxx037)
- Lipsitch M, Jha A, Simonsen L. 2016 Observational studies and the difficult quest for causality: lessons from vaccine effectiveness and impact studies. *Int. J. Epidemiol.* **45**, 2060–2074. (doi:10.1093/ije/dyw124)
- Jackson ML, Rothman KJ. 2015 Effects of imperfect test sensitivity and specificity on observational studies of influenza vaccine effectiveness. *Vaccine* **33**, 1313–1316. (doi:10.1016/j.vaccine.2015.01.069)
- Haber M, Lopman BA, Tate JE, Shi M, Parashar UD. 2018 A comparison of the test-negative and traditional case-control study designs with respect to the bias of estimates of rotavirus vaccine effectiveness. *Vaccine* **36**, 5071–5076. (doi:10.1016/j.vaccine.2018.06.072)
- Farrington CP. 1990 Quantifying misclassification bias in cohort studies of vaccine efficacy. *Stat. Med.* **9**, 1327–1337. (doi:10.1002/sim.4780091110)
- Sullivan SG, Tchetgen EJT, Cowling BJ. 2016 Theoretical basis of the test-negative study design for assessment of influenza vaccine effectiveness. *Am. J. Epidemiol.* **184**, 345–353. (doi:10.1093/aje/kww064)
- Halloran ME, Struchiner CJ, Longini IM. 1997 Study designs for evaluating different efficacy and effectiveness aspects of vaccines. *Am. J. Epidemiol.* **146**, 789–803. (doi:10.1093/oxfordjournals.aje.a009196)
- Wacholder S, Silverman DT, Mclaughlin JK, Mandel JS. 1992 Selection of controls in case-control studies: II. Types of controls. *Am. J. Epidemiol.* **135**, 1029–1041. (doi:10.1093/oxfordjournals.aje.a116397)
- Kotloff KL, Platts-Mills JA, Nasrin D, Roose A, Blackwelder WC, Levine MM. 2017 Global burden of diarrheal diseases among children in developing countries: incidence, aetiology, and insights from new molecular diagnostic techniques. *Vaccine* **35**, 6783–6789. (doi:10.1016/j.vaccine.2017.07.036)
- Panchalingam S *et al.* 2012 Diagnostic microbiologic methods in the GEMS-1 case/control study. *Clin. Infect. Dis.* **55**, S294–S302. (doi:10.1093/cid/cis754)
- Hammit LL, Murdoch DR, Scott JAG, Driscoll A, Karon RA, Levine OS, O'Brien KL. 2012 Specimen collection for the diagnosis of pediatric pneumonia. *Clin. Infect. Dis.* **54**, S132–S139. (doi:10.1093/cid/cir1068)
- Breslow NE. 1996 Statistics in epidemiology: the case-control study. *J. Am. Stat. Assoc.* **92**, 14–28. (doi:10.1080/01621459.1996.10476660)
- Lewnard JA, Lo NC, Arinaminpathy N, Frost I, Laxminarayan R. 2020 Childhood vaccines and antibiotic use in low- and middle-income countries. *Nature* **581**, 94–99. (doi:10.1038/s41586-020-2238-4)
- Lewnard JA, Givon-Lavi N, Dagan R. 2019 Dose-specific effectiveness of 7- and 13-valent pneumococcal conjugate vaccines against vaccine-serotype *Streptococcus pneumoniae* colonization in children. *Clin. Infect. Dis.* (doi:10.1093/cid/ciz1164)
- Gunn JS, Marshall JM, Baker S, Dongol S, Charles RC, Ryan ET. 2014 *Salmonella* chronic carriage: epidemiology, diagnosis, and gallbladder

- persistence. *Trends Microbiol.* **22**, 648–655. (doi:10.1016/j.tim.2014.06.007)
25. Apoola A, Herrero-Diaz M, FitzHugh E, Rajakumar R, Fakis A, Oakden J. 2011 A randomised controlled trial to assess pain with urethral swabs. *Sex. Transm. Infect.* **87**, 110–113. (doi:10.1136/sti.2010.042861)
  26. Rinta-Kokko H, Dagan R, Givon-Lavi N, Auranen K. 2009 Estimation of vaccine efficacy against acquisition of pneumococcal carriage. *Vaccine* **27**, 3831–3837. (doi:10.1016/j.vaccine.2009.04.009)
  27. Auranen K, Rinta-Kokko H, Goldblatt D, Nohynek H, O'Brien KL, Satzke C, Simell B, Tanskanen A, Käyhty H. 2013 Colonisation endpoints in *Streptococcus pneumoniae* vaccine trials. *Vaccine* **32**, 153–158. (doi:10.1016/j.vaccine.2013.08.061)
  28. Tse H, To KKW, Wen X, Chen H, Chan K-H, Tsoi H-W, Li IWS, Yuen K-Y. 2011 Clinical and virological factors associated with viremia in pandemic influenza A/H1N1/2009 virus infection. *PLoS ONE* **6**, e22534. (doi:10.1371/journal.pone.0022534)
  29. Choi SM, Xie H, Campbell AP, Kuypers J, Leisenring W, Boudreaux AA, Englund JA, Corey L, Boeckh M. 2012 Influenza viral RNA detection in blood as a marker to predict disease severity in hematopoietic cell transplant recipients. *J. Infect. Dis.* **206**, 1872–1877. (doi:10.1093/infdis/jis610)
  30. Torres JP *et al.* 2010 Respiratory syncytial virus (RSV) RNA loads in peripheral blood correlates with disease severity in mice. *Respir. Res.* **11**, 125. (doi:10.1186/1465-9921-11-125)
  31. Esposito S, Daleno C, Scala A, Castellazzi L, Terranova L, Sferazza Papa S, Longo MR, Pelucchi C, Principi N. 2014 Impact of rhinovirus nasopharyngeal viral load and viremia on severity of respiratory infections in children. *Eur. J. Clin. Microbiol. Infect. Dis.* **33**, 41–48. (doi:10.1007/s10096-013-1926-5)
  32. Park DE *et al.* 2017 Colonization density of the upper respiratory tract as a predictor of pneumonia: *Haemophilus influenzae*, *Moraxella catarrhalis*, *Staphylococcus aureus*, and *Pneumocystis jirovecii*. *Clin. Infect. Dis.* **64**, S328–S336. (doi:10.1093/cid/cix104)
  33. Murdoch DR *et al.* 2003 Evaluation of a PCR assay for detection of *Streptococcus pneumoniae* in respiratory and nonrespiratory samples from adults with community-acquired pneumonia. *J. Clin. Microbiol.* **41**, 63–66. (doi:10.1128/JCM.41.1.63-66.2003)
  34. Michelow IC, Lozano J, Olsen K, Goto C, Rollins NK, Ghaffar F, Violeta R-C, Leinonen M, McCracken GH. 2002 Diagnosis of *Streptococcus pneumoniae* lower respiratory infection in hospitalized children by culture, polymerase chain reaction, serological testing, and urinary antigen detection. *Clin. Infect. Dis.* **34**, e1–e11. (doi:10.1086/324358)
  35. Dangerfield B, Chung A, Webb B, Seville MT. 2014 Predictive value of methicillin-resistant *Staphylococcus aureus* (MRSA) nasal swab PCR assay for MRSA pneumonia. *Antimicrob. Agents Chemother.* **58**, 859–864. (doi:10.1128/AAC.01805-13)
  36. Chua H, Feng S, Lewnard JA, Sullivan SG, Blyth CC, Lipsitch M, Cowling BJ. 2020 The use of test-negative controls to monitor vaccine effectiveness. *Epidemiology* **31**, 43–64. (doi:10.1097/EDE.0000000000001116)
  37. Foppa IM, Haber M, Ferdinands JM, Shay DK. 2013 The case test-negative design for studies of the effectiveness of influenza vaccine. *Vaccine* **31**, 3104–3109. (doi:10.1016/j.vaccine.2013.04.026)
  38. Lewnard JA, Tedijanto C, Cowling BJ, Lipsitch M. 2018 Measurement of vaccine direct effects under the test-negative design. *Am. J. Epidemiol.* **187**, 2686–2697. (doi:10.1093/aje/kwy163)
  39. Mulholland K *et al.* 1997 Randomised trial of *Haemophilus influenzae* type-b tetanus protein conjugate for prevention of pneumonia and meningitis in Gambian infants. *Lancet* **349**, 1191–1197. (doi:10.1016/S0140-6736(96)09267-7)
  40. Gessner BD *et al.* 2005 Incidences of vaccine-preventable *Haemophilus influenzae* type b pneumonia and meningitis in Indonesian children: Hamlet-randomised vaccine-probe trial. *Lancet* **365**, 43–52. (doi:10.1016/S0140-6736(04)17664-2)
  41. Feiken DR, Scott JAG, Gessner BD. 2014 Use of vaccines as probes to define disease burden. *Lancet* **383**, 1762–1770. (doi:10.1016/S0140-6736(13)61682-7)
  42. Halloran ME, Struchiner CJ. 1995 Causal inference in infectious diseases. *Epidemiology* **6**, 142–151. (doi:10.1097/00001648-199503000-00010)
  43. Westreich D, Hudgens MG. 2016 Beware the test-negative design. *Am. J. Epidemiol.* **184**, 354–356. (doi:10.1093/aje/kww063)