

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

Joseph A. Lewnard

Correspondence to Dr. Joseph Lewnard, Division of Epidemiology and Biostatistics, School of Public Health, University of California, Berkeley, 2121 Berkeley Way, Berkeley, CA 94720 (e-mail: jlewnard@berkeley.edu); (510) 664-4050

Author affiliations: Division of Epidemiology and Biostatistics, School of Public Health, University of California, Berkeley, Berkeley, California

Funding: The author received no specific funding for this article.

ABSTRACT

The diagnosis of infectious disease syndromes such as fever, diarrhea, and pneumonia is complicated by the potential for shedding or carriage of putatively etiologic pathogens among individuals experiencing symptoms due to other causes. Symptomatic individuals among whom a pathogen is detected, but whose symptoms are caused by other factors, may be misclassified by diagnostic criteria based on pathogen detection. Case-control studies are commonly undertaken to estimate vaccine effectiveness, and present an opportunity to compare pathogen detection among individuals with and without clinically-relevant symptoms. Considering this study context, we derive 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 etiologic fractions. We also derive biases affecting vaccine effectiveness estimates under the test-negative design, an alternative approach using similar data. Our results provide strategies for estimation of pathogen-specific vaccine effectiveness in the absence a diagnostic gold standard. These approaches can inform the design and analysis of studies addressing numerous pathogens and vaccines.

KEYWORDS

Case-control studies; Infectious disease; Misclassification; Test-negative design; Vaccine effectiveness

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 etiology. Such is the case for pathogens that may be shed or carried in a subclinical state as part of their natural history. For instance, viruses and bacteria that cause pneumonia may be found in respiratory tract specimens from individuals experiencing acute symptoms as well as individuals without clinically-apparent illness, especially among children (1). Similar circumstances arise in the natural history of pathogens that cause diarrhea (2) and acute febrile illness (3). 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 (4).

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. Recently, large-scale studies of pneumonia (5) and diarrhea (6) etiology 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 center on a comparison of the prevalence of each pathogen among symptomatic cases and asymptomatic controls (7–9). While similar case-control frameworks may present an

opportunity to correct for misclassification when estimating vaccine effectiveness (VE; (10–14)), 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 (15). These effects address the biologic 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.

STUDY DESIGN

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 (16). Assume that all or a proportion of individuals enrolled (including both cases and controls) are tested to determine presence or 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 etiology.

Examples

Case-control studies involving the detection of enteropathogens from diarrhea cases and asymptomatic controls, or the detection of respiratory viruses and bacteria in the upper respiratory tract of case with and controls without acute respiratory symptoms, present ideal environments for application of the approaches detailed below. Studies of diarrhea-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 stool specimens regardless of the presence of symptoms, and the detection of a pathogen in diarrheal stool is not always a clear indication of its etiologic significance (8,17,18). Similar concerns arise with detection of arboviruses, *Rickettsia*, and malaria parasites among individuals experiencing acute febrile illness in endemic settings (3). Although these challenges may be mitigated for certain pathogens by the use of quantitative molecular diagnostic tests (4), such testing remains relatively uncommon in clinical practice.

In studies of pediatric pneumonia, inaccessibility of the site of infection (i.e., the lung) presents a unique challenge where the approaches we consider below may offer further 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 etiologic 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 (19).

NOTATION AND THEORETICAL FRAMEWORK

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. Define X_i as an indicator of the presence ($X_i = 1$) or absence ($X_i = 0$) of the pathogen of interest, and define Z_i as an indicator of the individual's vaccine status as vaccinated (exposed, $Z_i = 1$) or unvaccinated (unexposed, $Z_i = 0$; **Table 1**).

We next consider epidemiologic parameters describing the prevalence and natural history of the vaccine-preventable pathogen of interest among cases and controls(**Table 2**). Define the prevalence of carriage or shedding of the pathogen in the absence of disease and vaccine-derived immunity as $\pi = \Pr(X_i = 1|Y_i = 0, Z_i = 0)$. Define $\theta_S\pi$ as the prevalence of shedding or carriage of 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 due 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 θ_S with respect to vaccine effects on the acquisition and duration of shedding in a later section.

Define ρ as the risk for the pathogen of interest, once acquired, to progress to disease in the absence of vaccine-conferred protection. Define θ_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 ω to represent the risk (measured over the same time interval as ρ) 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 carriage or shedding of the pathogen of interest. We derive ρ and ω with respect to rate parameters in a later section to assess the relation of θ_P to the hazard ratio of progression of the pathogen of interest to symptomatic disease, given vaccination.

Accounting for disease attributable to the pathogen of interest and all other pathogens, we have

$$\begin{aligned}\Pr(Y_i = 1|X_i = 1, Z_i = 0) &= \omega + \rho \\ \Pr(Y_i = 1|X_i = 1, Z_i = 1) &= \omega + \theta\rho \\ \Pr(Y_i = 1|X_i = 0, Z_i = 0) &= \Pr(Y_i = 1|X_i = 0, Z_i = 1) = \omega\end{aligned}\tag{1a-c}$$

Note that in a case-control study, enrollment 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 which do not require prospective observation of the probabilities presented in Eq. 1.

VACCINE DIRECT EFFECT AGAINST SHEDDING OR CARRIAGE OF THE PATHOGEN OF INTEREST (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 of the pathogen of interest among controls, given vaccination, provides the most straightforward basis for estimating VE_S :

$$RR_S = \frac{\Pr(X_i = 1|Y_i = 0, Z_i = 1)}{\Pr(X_i = 1|Y_i = 0, Z_i = 0)} = \frac{\theta_S \pi}{\pi} = \theta_S = 1 - VE_S. \quad (2)$$

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 (20). Such an approach may be advantageous 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 (21,22)), or when favored by other design considerations. This odds ratio (OR) can be expressed as

$$OR_S = \frac{[\Pr(Z_i = 1, X_i = 1|Y_i = 0)][\Pr(Z_i = 0|Y_i = 0)]}{[\Pr(Z_i = 0, X_i = 1|Y_i = 0)][\Pr(Z_i = 1|Y_i = 0)]}. \quad (3)$$

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,

$$\Pr(Z_i = 1, X_i = 1|Y_i = 0) = \Pr(X_i = 1|Z_i = 1, Y_i = 0) \Pr(Z_i = 1|Y_i = 0) = \theta_S \pi v \quad (4a)$$

and

$$\Pr(Z_i = 0, X_i = 1 | Y_i = 0) = \Pr(X_i = 1 | Z_i = 0, Y_i = 0) \Pr(Z_i = 0 | Y_i = 0) = \pi(1 - v). \quad (4b)$$

Substituting into Eq. 3,

$$\text{OR}_S = \frac{[\theta_S \pi v] [(1 - v)]}{[\pi(1 - v)][v]} = \theta_S = 1 - \text{VE}_S. \quad (5)$$

VACCINE DIRECT EFFECT AGAINST PROGRESSION OF THE PATHOGEN OF INTEREST TO DISEASE FOLLOWING ACQUISITION (VE_p)

Estimator

Here we address the vaccine-attributable reduction in susceptibility to disease, given a pathogen has overcome a host's vaccine-derived immunity against acquisition of the carriage or shedding state. To estimate this effect, consider first the OR of detection of the pathogen of interest, given symptoms. We expect this OR to exceed one if acquisition of the pathogen increases individuals' risk of experiencing symptoms (i.e., $\rho > 0$; (7,8)). Among the unvaccinated,

$$\text{OR}_p^0 = \frac{\Pr(X_i = 1 | Y_i = 1, Z_i = 0) \Pr(X_i = 0 | Y_i = 0, Z_i = 0)}{\Pr(X_i = 0 | Y_i = 1, Z_i = 0) \Pr(X_i = 1 | Y_i = 0, Z_i = 0)}, \quad (7a)$$

while among the vaccinated,

$$\text{OR}_p^1 = \frac{\Pr(X_i = 1 | Y_i = 1, Z_i = 1) \Pr(X_i = 0 | Y_i = 0, Z_i = 1)}{\Pr(X_i = 0 | Y_i = 1, Z_i = 1) \Pr(X_i = 1 | Y_i = 0, Z_i = 1)}. \quad (7b)$$

We may expect that $OR_P^1 < OR_P^0$ if vaccination confers protection against progression of the pathogen from shedding to disease, thereby attenuating the association between pathogen detection and symptoms among the vaccinated. Of the terms appearing in Eq. 7, 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 Eq. 2. From the progression parameters ρ , ω , and θ_p , we may further define

$$\Pr(X_i = 1|Y_i = 1, Z_i = 0) = \frac{\pi(\rho + \omega)}{\pi\rho + \omega} \tag{8a}$$

and

$$\Pr(X_i = 1|Y_i = 1, Z_i = 1) = \frac{\theta_S\pi(\theta_p\rho + \omega)}{(\theta_S\pi\theta_p\rho + \omega)}. \tag{8b}$$

As the complements of Eqs. 8a and 8b, respectively,

$$\Pr(X_i = 0|Y_i = 1, Z_i = 0) = \frac{(1 - \pi)\omega}{\pi\rho + \omega} \tag{8c}$$

and

$$\Pr(X_i = 0|Y_i = 1, Z_i = 1) = \frac{(1 - \theta_S\pi)\omega}{\theta_S\pi\theta_p\rho + \omega}, \tag{8d}$$

Substituting terms from Eqs. 2 and 8 into Eq. 7, we have

$$OR_P^0 = \frac{[\pi(\rho + \omega)][(1 - \pi)]}{[\omega(1 - \pi)][\pi]} = \frac{\rho + \omega}{\omega} \tag{9a}$$

and

$$\text{OR}_p^1 = \frac{[\theta_S \pi (\theta_P \rho + \omega)] [(1 - \theta_S \pi)]}{[(1 - \theta_S \pi) \omega] [\theta_S \pi]} = \frac{\theta_P \rho + \omega}{\omega}. \quad (9b)$$

Here, $\text{OR}_p^1 < \text{OR}_p^0$ for $\theta_P < 1$ and $\rho > 0$, $\omega > 0$. Rearranging the terms from Eq. 9 reveals $\rho = \omega(\text{OR}_p^0 - 1)$ and $\theta_P \rho = \omega(\text{OR}_p^1 - 1)$. Substituting for θ_P ,

$$\text{VE}_p = 1 - \theta_P = 1 - \frac{\text{OR}_p^1 - 1}{\text{OR}_p^0 - 1}. \quad (10)$$

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

Estimation considerations

Estimation of the vaccine direct effect against progression requires division of the parameters OR_p^1 and OR_p^0 . When these quantities are estimated independently, variance of the ratio $(\text{OR}_p^1 - 1)/(\text{OR}_p^0 - 1)$ may be overstated due to the failure to account for the covariance of these terms; further, because we envision an observational study design, it may be necessary to control for confounding factors that independently predict individuals' likelihood of pathogen detection, vaccination, and disease. These concerns may be alleviated by estimating OR_p^0 and OR_p^1 via the logistic regression model

$$\log\left(\frac{Y_i}{1 - Y_i}\right) = \beta_0 + \beta_1 X_i + \beta_2 Z_i + \beta_3 X_i Z_i + \sum_k \delta_k W_{i,k} + \epsilon_i,$$

where $\epsilon_i \sim N(0, \sigma^2)$. Here, we may interpret $OR_p^0 = \exp(\beta_1)$ and $OR_p^1 = \exp(\beta_1 + \beta_3)$ as adjusted estimates of the association between presence of the pathogen and disease risk. The parameters δ_k convey the effects of confounding factors W_k , and β_2 conveys any difference in disease risk among the vaccinated and unvaccinated not mediated by protection against the pathogen of interest. Under this framework, we may express the vaccine direct effect against progression in terms of the regression parameters

$$VE_p = 1 - \frac{\exp(\beta_1 + \beta_3) - 1}{\exp(\beta_1) - 1}. \quad (11)$$

VACCINE DIRECT EFFECT AGAINST DISEASE ATTRIBUTABLE TO THE PATHOGEN OF INTEREST (VE_D)

We last consider the cumulative extent of vaccine-conferred protection resulting from prevention of shedding or carriage of the pathogen of interest, 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,

$$VE_D = 1 - \frac{\Pr(\text{Pathogen attributable disease} | \text{Vaccinated})}{\Pr(\text{Pathogen attributable disease} | \text{Unvaccinated})} = 1 - \frac{\theta_S \pi \theta_P \rho}{\pi \rho} = 1 - \theta_S \theta_P, \quad (12)$$

which may thus be estimated as

$$VE_D = 1 - RR_S \times \frac{OR_p^1 - 1}{OR_p^0 - 1} \quad (13a)$$

or

$$VE_D = 1 - OR_S \times \frac{OR_p^1 - 1}{OR_p^0 - 1}, \quad (13b)$$

as defined above.

INTERPRETATION OF RISKS AND RATES

Pathogen shedding or carriage parameters

Interpretation of pathogen prevalence measures presents a challenge in studies of VE against endpoints such as asymptomatic shedding or carriage of a pathogen due to the dynamic nature of transmission (15,23,24). Here we relate $\theta_S = 1 - VE_S$ to vaccine direct effects on the rates of pathogen acquisition and clearance.

Define λ and γ as the rates at which unvaccinated individuals acquire and clear the pathogen of interest, respectively. Taking π 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 relative rates of acquisition and clearance of the pathogen, given vaccination, we have $a\lambda(1 - \theta_S\pi) = b\gamma\theta_S\pi$. Substituting for π , we obtain

$$\theta_S = \frac{a(\lambda + \gamma)}{a\lambda + b\gamma}. \quad (14)$$

If the rate of clearance of the pathogen far exceeds the rate of acquisition ($\gamma \gg \lambda$), we have $\theta_S \cong \lim_{\lambda/\gamma \rightarrow 0} (a(\lambda + \gamma)/(a\lambda + b\gamma)) = a/b$. Thus, the prevalence ratio θ_S among controls (as estimated by RR_S or OR_S) may be interpretable as the hazard ratio of acquisition if $b = 1$, or as the inverse of the hazard ratio of clearance if $a = 1$.

Disease progression parameters

We may relate the risk parameters ρ , ω , and θ_P to epidemiologic rates under a similar intuition. Define r_ρ as the rate of progression of the pathogen of interest to disease, following acquisition by an unvaccinated individual, and define r_ω as the rate of onset of disease due to all other causes. Here we may use a competing risks framework to define ρ and ω 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

$$\rho = \int_0^{\infty} f(t|r_\rho)(1 - F(t|r_\omega + \gamma))dt \tag{15a}$$

and

$$\omega = \int_0^{\infty} f(t|r_\omega)(1 - F(t|r_\rho + \gamma))dt. \tag{15b}$$

Next defining c as the hazard ratio of progression of the pathogen of interest to disease following acquisition, given vaccination, we have

$$\theta_P \rho = \int_0^{\infty} f(t|cr_\rho)(1 - F(t|r_\omega + \gamma))dt.$$

(15c)

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

$$\theta_P = \frac{c(r_\rho + r_\omega + \gamma)}{cr_\rho + r_\omega + \gamma}. \quad (16)$$

If the rate of progression of the pathogen of interest is well below the rate of disease onset due to all other causes ($r_\omega \gg r_\rho$), or the rate of clearance of the pathogen ($\gamma \gg r_\rho$), we have $\theta_P \cong$

$\lim_{r_\rho / (r_\omega + \gamma) \rightarrow 0} (c(r_\rho + r_\omega + \gamma) / (cr_\rho + r_\omega + \gamma)) = c$. Thus, for diseases that are uncommon, $OR_P =$

θ_P may be interpretable as the hazard ratio of disease progression, following acquisition, for a vaccinated versus unvaccinated individual.

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 (25–28).

In the event that we may exclude alternative etiologies among cases in whom the pathogen is detected, Eqs. 8a and 8b may be rewritten as

$$\Pr(X_i = 1|Y_i = 1, Z_i = 0) = \frac{\pi\rho}{\pi\rho + (1 - \pi)\omega} \quad (17a)$$

and

$$\Pr(X_i = 1|Y_i = 1, Z_i = 1) = \frac{\theta_S\pi\theta_P\rho}{\theta_S\pi\theta_P\rho + (1 - \theta_S\pi)\omega}, \quad (17b)$$

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. Here,

$$\text{OR}_P^0 = \frac{[\pi\rho][(1 - \pi)]}{[(1 - \pi)\omega][\pi]} = \frac{\rho}{\omega} \quad (18a)$$

and

$$\text{OR}_P^1 = \frac{[\theta_S\pi\theta_P\rho][(1 - \theta_S\pi)]}{[(1 - \theta_S\pi)\omega][\theta_S\pi]} = \frac{\theta_P\rho}{\omega}, \quad (18b)$$

so that

$$\text{VE}_P = 1 - \theta_P = 1 - \frac{\text{OR}_P^1}{\text{OR}_P^0}, \quad (19)$$

in contrast to Eq.10.

Additionally, this special case allows VE_D to be estimated directly by comparing the odds of prior vaccination among cases with the pathogen detected versus controls among whom the pathogen is not detected. We show $\Pr(Z_i = z|X_i = 1, Y_i = 0)$ in Eq. 4. Similarly,

$$\Pr(Z_i = 1|X_i = 1, Y_i = 1) = \frac{\Pr(X_i = 1, Y_i = 1|Z_i = 1) \Pr(Z_i = 1)}{\Pr(X_i = 1, Y_i = 1)} = \frac{\theta_S \theta_P v}{\theta_S \theta_P v + (1 - v)} \quad (20a)$$

and its complement

$$\Pr(Z_i = 0|X_i = 1, Y_i = 1) = \frac{1 - v}{\theta_S \theta_P v + (1 - v)}, \quad (20b)$$

under the circumstance that the pathogen, if detected in a symptomatic individual, is the true cause of disease. Taken together,

$$\begin{aligned} VE_D &= 1 - \frac{\Pr(Z_i = 1|X_i = 1, Y_i = 1) \Pr(Z_i = 0|X_i = 0, Y_i = 0)}{\Pr(Z_i = 0|X_i = 1, Y_i = 1) \Pr(Z_i = 1|X_i = 0, Y_i = 0)} \\ &= 1 - \frac{[\theta_S \theta_P v][(1 - \pi)(1 - v)]}{[(1 - v)][(1 - \pi)v]} = 1 - \theta_S \theta_P. \end{aligned} \quad (21)$$

BIAS UNDER IMPERFECT TEST SENSITIVITY

For certain conditions, 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 (4,29).

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 (30–32).

Take $D_i = 1$ and $D_i = 0$ to indicate detection or non-detection of the pathogen of interest. To account for test sensitivity in the formulations introduced above, we may replace X_i with D_i in our previous equations, so that

$$RR_S = \frac{\Pr(D_i = 1|Y_i = 0, Z_i = 1)}{\Pr(D_i = 1|Y_i = 0, Z_i = 0)} \quad (22a)$$

and

$$OR_S = \frac{[\Pr(Z_i = 1, D_i = 1|Y_i = 0)][\Pr(Z_i = 0|Y_i = 0)]}{[\Pr(Z_i = 0, D_i = 1|Y_i = 0)][\Pr(Z_i = 1|Y_i = 0)]} \quad (22b)$$

for estimation of VE_S , while

$$OR_P^0 = \frac{\Pr(D_i = 1|Y_i = 1, Z_i = 0) \Pr(D_i = 0|Y_i = 0, Z_i = 0)}{\Pr(D_i = 0|Y_i = 1, Z_i = 0) \Pr(D_i = 1|Y_i = 0, Z_i = 0)} \quad (23a)$$

and

$$OR_P^1 = \frac{\Pr(D_i = 1|Y_i = 1, Z_i = 1) \Pr(D_i = 0|Y_i = 0, Z_i = 1)}{\Pr(D_i = 0|Y_i = 1, Z_i = 1) \Pr(D_i = 1|Y_i = 0, Z_i = 1)} \quad (23b)$$

for estimation of VE_D .

Define α_1 and α_0 as the probability of detecting the pathogen of interest, given it is present, for individuals who are and are not 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 Eq. 17 reveals no effect on RR_S and OR_S , as α_0 cancels in the numerator and denominator.

Next considering OR_P^0 , we have

$$\Pr(D_i = 1|Y_i = 1, Z_i = 0) = \frac{\Pr(Y_i = 1, Z_i = 0|X_i = 1, D_i = 1) \Pr(X_i = 1, D_i = 1)}{\Pr(Y_i = 1, Z_i = 0)}, \quad (24a)$$

since $\Pr(X_i = 0, D_i = 1) = 0$. Based on the parameter definitions above,

$$\Pr(D_i = 1|Y_i = 1, Z_i = 0) = \frac{\pi(\alpha_1\rho + \alpha_0\omega)}{\pi\rho + \omega} \quad (24b)$$

and

$$\Pr(D_i = 0|Y_i = 1, Z_i = 0) = \frac{\pi\rho(1 - \alpha_1) + \omega(1 - \alpha_0\pi)}{\pi\rho + \omega}. \quad (24c)$$

Similarly, among the vaccinated,

$$\Pr(D_i = 1|Y_i = 1, Z_i = 1) = \frac{\theta_S\pi(\alpha_1\theta_P\rho + \alpha_0\omega)}{\theta_S\pi\theta_P\rho + \omega} \quad (25a)$$

and

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

(25b)

Substituting into Eq. 18, we have

$$OR_P^0 = \frac{[\pi(\alpha_1\rho + \alpha_0\omega)][1 - \alpha_0\pi]}{[\pi\rho(1 - \alpha_1) + \omega(1 - \alpha_0\pi)][\alpha_0\pi]} = \frac{(\alpha_1\rho + \alpha_0\omega)(1 - \alpha_0\pi)}{\alpha_0[\pi\rho(1 - \alpha_1) + \omega(1 - \alpha_0\pi)]}$$

(26a)

and

$$OR_P^1 = \frac{[\theta_S\pi(\alpha_1\theta_P\rho + \alpha_0\omega)][1 - \alpha_0\theta_S\pi]}{[\theta_S\pi\theta_P\rho(1 - \alpha_1) + \omega(1 - \alpha_0\theta_S\pi)][\alpha_0\theta_S\pi]} = \frac{(\alpha_1\theta_P\rho + \alpha_0\omega)(1 - \alpha_0\theta_S\pi)}{\alpha_0[\theta_S\pi\theta_P\rho(1 - \alpha_1) + \omega(1 - \alpha_0\theta_S\pi)]}$$

(26b)

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 parameterizations. The extent of bias increases with higher prevalence of the pathogen among persons without symptoms (π), with higher rates of progression of the pathogen to disease (ρ/ω), and with lower sensitivity for pathogen detection in disease (α_1). Interestingly, bias also worsens as α_0 approaches 1.

The nature and extent of bias further depends upon whether vaccination confers protection against pathogen shedding or carriage (θ_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 $VE_P < 0\%$. In 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, test sensitivity is very high in the context of subclinical shedding or carriage and low among those experiencing disease, and true protection against disease progression is very weak.

BIAS UNDER TEST-NEGATIVE DESIGN APPROACHES

Protection against disease

The test-negative design (TND) has come into routine use in studies of VE (33). These studies 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}). Case-control studies such as we consider here may resemble TND studies in that receipt of a test is independent of the likelihood of detecting the pathogen of interest (14,34,35). It has been previously recognized that shedding or carriage of a pathogen among individuals experiencing symptoms due to other causes may influence TND estimates, although the extent of resulting bias is not well understood (10,12,14). 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

$$\widehat{VE}_D = 1 - OR_D^{TND} = 1 - \frac{\Pr(Z_i = 1|X_i = 1, Y_i = 1) \Pr(Z_i = 0|X_i = 0, Y_i = 1)}{\Pr(Z_i = 0|X_i = 1, Y_i = 1) \Pr(Z_i = 1|X_i = 0, Y_i = 1)}. \quad (27)$$

Here we may define

$$\begin{aligned} \Pr(Z_i = 1|X_i = 1, Y_i = 1) &= \frac{\Pr(X_i = 1, Y_i = 1|Z_i = 1) \Pr(Z_i = 1)}{\Pr(X_i = 1, Y_i = 1)} \\ &= \frac{\theta_S(\theta_P\rho + \omega)v}{\theta_S(\theta_P\rho + \omega)v + (\rho + \omega)(1 - v)} \end{aligned} \quad (28a)$$

and

$$\Pr(Z_i = 0|X_i = 1, Y_i = 1) = \frac{(\rho + \omega)(1 - v)}{\theta_S(\theta_P\rho + \omega)v + (\rho + \omega)(1 - v)}. \quad (28b)$$

Similarly, we have

$$\Pr(Z_i = 1|X_i = 0, Y_i = 1) = \frac{(1 - \theta_S\pi)v}{(1 - \theta_S\pi)v + (1 - \pi)(1 - v)} \quad (29a)$$

and

$$\Pr(Z_i = 0|X_i = 0, Y_i = 1) = \frac{(1 - \pi)(1 - v)}{(1 - \theta_S\pi)v + (1 - \pi)(1 - v)}. \quad (29b)$$

Substituting Eqs. 28 and 29 into Eq. 27, we obtain

$$OR_D^{TND} = \frac{[\theta_S(\theta_P\rho + \omega)v][(1 - \pi)(1 - v)]}{[(\rho + \omega)(1 - v)][(1 - \theta_S\pi)v]} = \frac{\theta_S(\theta_P\rho + \omega)(1 - \pi)}{(\rho + \omega)(1 - \theta_S\pi)}. \quad (30)$$

This estimator approaches the true vaccine direct effect against acquisition and progression of the pathogen of interest ($VE_D = 1 - \theta_S\theta_P$) as π and ω/ρ approach zero, and as θ_S and θ_D approach zero (resulting in near-100% protection). We illustrate the quantitative extent of bias under differing conditions in **Figure 2**.

Bias in the absence of diagnostic uncertainty

We also identify bias under the TND framework under the circumstances addressed in section 7, where detection of the pathogen from cases can be assumed to indicate true disease etiology.

Following Eq. 20, in this instance we may consider

$$\begin{aligned} \Pr(Z_i = 1|X_i = 0, Y_i = 1) &= \frac{\Pr(X_i = 0, Y_i = 1|Z_i = 1) \Pr(Z_i = 1)}{\Pr(X_i = 0, Y_i = 1)} \\ &= \frac{(1 - \theta_S\pi)v}{(1 - \theta_S\pi)v + (1 - \pi)(1 - v)} \end{aligned} \tag{31a}$$

and

$$\Pr(Z_i = 0|X_i = 0, Y_i = 1) = \frac{(1 - \pi)(1 - v)}{(1 - \theta_S\pi)v + (1 - \pi)(1 - v)}; \tag{31b}$$

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 Eq. 20, we have

$$OR_D^{TND} = \frac{[\theta_S\theta_P v][(1 - \pi)(1 - v)]}{[(1 - v)][(1 - \theta_S\pi)v]} = \frac{\theta_S\theta_P(1 - \pi)}{1 - \theta_S\pi}, \tag{32}$$

which converges to $\theta_S\theta_P$ as θ_S , θ_P , or π approach zero (**Figure 2**). Otherwise, this approach will overestimate the magnitude of the true vaccine direct effect (i.e., $OR_D^{TND} < \theta_S\theta_P$).

Protection against carriage or shedding

Our mathematical framework reveals similar biases affecting estimation of VE_S under TND-like approaches. Although TND studies typically enroll 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 pathogen of interest. Here, estimates would be expected to indicate vaccine effectiveness against acquisitions that result only in subclinical shedding or carriage (10,33,35).

Consider the odds ratio of vaccination among control individuals who test positive or negative for the pathogen of interest:

$$OR_S^{TND} = \frac{[\Pr(Z_i = 1, X_i = 1|Y_i = 0)][\Pr(Z_i = 0, X_i = 0|Y_i = 0)]}{[\Pr(Z_i = 0, X_i = 1|Y_i = 0)][\Pr(Z_i = 1, X_i = 0|Y_i = 0)]} \quad (33)$$

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 Eq. 4. We may similarly derive

$$\Pr(Z_i = 1, X_i = 0|Y_i = 0) = \Pr(X_i = 0|Z_i = 1, Y_i = 0) \Pr(Z_i = 1|Y_i = 0) = (1 - \theta_S\pi)v \quad (34a)$$

and

$$\Pr(Z_i = 0, X_i = 0|Y_i = 0) = \Pr(X_i = 0|Z_i = 0, Y_i = 0) \Pr(Z_i = 0|Y_i = 0) = (1 - \pi)(1 - v). \quad (34b)$$

Substituting into Eq. 33, we have

$$\text{OR}_S^{\text{TND}} = \frac{[\theta_S \pi v][(1 - \pi)(1 - v)]}{[\pi(1 - v)][(1 - \theta_S \pi)v]} = \frac{\theta_S(1 - \pi)}{1 - \theta_S \pi}, \quad (35)$$

revealing the same form of bias inherent to Eq. 32, above.

ATTRIBUTABLE FRACTION ESTIMATION

Data from case-control studies such as we describe 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 (36,37), 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

$$\text{OR}_A = \frac{\Pr(Z_i = 1|Y_i = 1) \Pr(Z_i = 0|Y_i = 0)}{\Pr(Z_i = 0|Y_i = 1) \Pr(Z_i = 1|Y_i = 0)}. \quad (36)$$

As in Eq. 3, we have $\Pr(Z_i = 1|Y_i = 0) = v$ and $\Pr(Z_i = 0|Y_i = 0) = 1 - v$, while

$$\Pr(Z_i = 1|Y_i = 1) = \frac{\Pr(Y_i = 1|Z_i = 1) \Pr(Z_i = 1)}{\Pr(Y_i = 1)} = \frac{(\theta_S \pi \theta_P \rho + \omega)v}{(\theta_S \pi \theta_P \rho + \omega)v + (\pi \rho + \omega)(1 - v)} \quad (37a)$$

and

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

(37b)

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

$$OR_A = \frac{[(\theta_S \pi \theta_P \rho + \omega)v][1 - v]}{[(\pi \rho + \omega)(1 - v)][v]} = \frac{\theta_S \pi \theta_P \rho + \omega}{\pi \rho + \omega}. \quad (38)$$

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 Eq. 38, we obtain the risk (or rate) ratio of disease attributable to the pathogen of interest, versus other causes, as

$$\frac{\pi \rho}{\omega} = \frac{1 - OR_A}{OR_A - \theta_S \theta_P}. \quad (39a)$$

where $1 - VE_D$ (Eq. 13) may provide an unbiased input for $\theta_S \theta_P$. Thus, the proportion of disease attributable to the pathogen of interest, absent vaccine-derived protection, is

$$\frac{\pi \rho}{\pi \rho + \omega} = \frac{1 - OR_A}{1 - \theta_S \theta_C}. \quad (39b)$$

SUMMARY

Microbiologic diagnosis of certain infectious diseases may be complicated by the prevalence of putative etiologic pathogens among individuals whose symptoms owe 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 likely 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 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 vaccine effectiveness 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 vaccine effectiveness, 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 (35,38). Nonetheless, our findings add to a growing list of concerns about uses of the TND for estimation of vaccine effectiveness (10,14,39). Control groups without symptoms potentially attributable to the pathogen of interest may be of value to correct for biases arising in conventional TND studies.

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 vaccine effectiveness against pathogen-specific endpoints without a diagnostic gold standard.

REFERENCES

1. Levine OS, O'Brien KL, Deloria-Knoll M, et al. The pneumonia etiology research for child health project: A 21st century childhood pneumonia etiology study. *Clin Infect Dis.* 2012;54:S93–S101.
2. Kotloff KL, Blackwelder WC, Nasrin D, et al. The Global Enteric Multicenter Study (GEMS) of diarrheal disease in infants and young children in developing countries: Epidemiologic and clinical methods of the case/control study. *Clin Infect Dis.* 2012;55:S232–S245.
3. Tam PYI, Obaro SK, Storch G. Challenges in the etiology and diagnosis of acute febrile illness in children in low- and middle- income countries. *J Pediatric Infect Dis Soc.* 2016;5:190–205.
4. Liu J, Kabir F, Manneh J, et al. Development and assessment of molecular diagnostic tests for 15 enteropathogens causing childhood diarrhoea: A multicentre study. *Lancet Infect Dis.* 2014;14:716–724.

5. O'Brien KL, Baggett HC, Brooks WA, et al. Causes of severe pneumonia requiring hospital admission in children without HIV infection from Africa and Asia: the PERCH multi-country case-control study. *Lancet*. 2019;394:757–779.
6. Kotloff KL, Nataro JP, Blackwelder WC, et al. 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*. 2013;382(9888):209–222.
7. Bruzzi P, Green SB, Byar DP, et al. Estimating the population attributable risk for multiple risk factors using case-control data. *Am J Epidemiol*. 1985;122:904–914.
8. Blackwelder WC, Biswas K, Wu Y, et al. Statistical methods in the Global Enteric Multicenter Study (GEMS). *Clin Infect Dis*. 2012;55:S246–S253.
9. Wu Z, Deloria-Knoll M, Zeger SL. Nested partially latent class models for dependent binary data; Estimating disease etiology. *Biostatistics*. 2017;18:200–213.
10. Lipsitch M, Jha A, Simonsen L. Observational studies and the difficult quest for causality: Lessons from vaccine effectiveness and impact studies. *Int J Epidemiol*. 2016;45(6):2060–2074.
11. Jackson ML, Rothman KJ. Effects of imperfect test sensitivity and specificity on observational studies of influenza vaccine effectiveness. *Vaccine*. 2015;33(11):1313–1316.
12. Haber M, Lopman BA, Tate JE, et al. A comparison of the test-negative and traditional case-control study designs with respect to the bias of estimates of rotavirus vaccine effectiveness. *Vaccine*. 2018;36(33):5071–5076.
13. Farrington CP. Quantifying misclassification bias in cohort studies of vaccine efficacy.

- Stat Med. 1990;9:1327–1337.
14. Sullivan SG, Tchetgen EJT, Cowling BJ. Theoretical Basis of the Test-Negative Study Design for Assessment of Influenza Vaccine Effectiveness. *Am J Epidemiol.* 2016;184(5):345–353.
 15. Halloran ME, Struchiner CJ, Longini IM. Study designs for evaluating different efficacy and effectiveness aspects of vaccines. *Am J Epidemiol.* 1997;146(10):789–803.
 16. Wacholder S, Silverman DT, Mclaughlin JK, et al. Selection of controls in case-control studies: II. Types of controls. *Am J Epidemiol.* 1992;135(9):1029–1041.
 17. Kotloff KL, Platts-Mills JA, Nasrin D, et al. Global burden of diarrheal diseases among children in developing countries: Incidence, etiology, and insights from new molecular diagnostic techniques. *Vaccine.* 2017;35:6783–6789.
 18. Panchalingam S, Antonio M, Hossain A, et al. Diagnostic microbiologic methods in the GEMS-1 case/control study. *Clin Infect Dis.* 2012;55:S294–S302.
 19. Hammitt LL, Murdoch DR, Scott JAG, et al. Specimen collection for the diagnosis of pediatric pneumonia. *Clin Infect Dis.* 2012;54:S132–S139.
 20. Lewnard JA, Givon-Lavi N, Dagan R. Dose-specific effectiveness of 7- and 13-valent pneumococcal conjugate vaccines against vaccine-serotype *Streptococcus pneumoniae* colonization in children. *Clin Infect Dis.* 2019; doi:10.1093/cid/ciz1164.
 21. Gunn JS, Marshall JM, Baker S, et al. Salmonella chronic carriage: Epidemiology, diagnosis, and gallbladder persistence. *Trends Microbiol.* 2014;22:648–655.
 22. Apoola A, Herrero-Diaz M, FitzHugh E, et al. A randomised controlled trial to assess pain with urethral swabs. *Sex Transm Infect.* 2011;87:110–113.
 23. Rinta-Kokko H, Dagan R, Givon-Lavi N, et al. Estimation of vaccine efficacy against

- acquisition of pneumococcal carriage. *Vaccine*. 2009;27(29):3831–3837.
24. Auranen K, Rinta-Kokko H, Goldblatt D, et al. Colonisation endpoints in *Streptococcus pneumoniae* vaccine trials. *Vaccine*. 2013;32:153–158.
25. Tse H, To KKW, Wen X, et al. Clinical and virological factors associated with viremia in pandemic influenza A/H1N1/2009 virus infection. *PLoS One*. 2011;6:e22534.
26. Choi SM, Xie H, Campbell AP, et al. Influenza viral RNA detection in blood as a marker to predict disease severity in hematopoietic cell transplant recipients. *J Infect Dis*. 2012;206:1872–1877.
27. Torres JP, Gomez AM, Khokhar S, et al. Respiratory Syncytial Virus (RSV) RNA loads in peripheral blood correlates with disease severity in mice. *Respir Res*. 2010;11:125.
28. Esposito S, Daleno C, Scala A, et al. Impact of rhinovirus nasopharyngeal viral load and viremia on severity of respiratory infections in children. *Eur J Clin Microbiol Infect Dis*. 2014;33:41–48.
29. Park DE, Baggett HC, Howie SRC, et al. Colonization density of the upper respiratory tract as a predictor of pneumonia: *Haemophilus influenzae*, *Moraxella catarrhalis*, *Staphylococcus aureus*, and *Pneumocystis jirovecii*. *Clin Infect Dis*. 2017;64:S328–S336.
30. Murdoch DR, Anderson TP, Beynon KA, et al. Evaluation of a PCR assay for detection of *Streptococcus pneumoniae* in respiratory and nonrespiratory samples from adults with community-acquired pneumonia. *J Clin Microbiol*. 2003;41:63–66.
31. Michelow IC, Lozano J, Olsen K, et al. Diagnosis of *Streptococcus pneumoniae* lower respiratory infection in hospitalized children by culture, polymerase chain reaction, serological testing, and urinary antigen detection. *Clin Infect Dis*. 2002;34:e1–e11.
32. Dangerfield B, Chung A, Webb B, et al. Predictive value of methicillin-resistant

staphylococcus aureus (MRSA) nasal swab PCR assay for MRSA pneumonia.

Antimicrob Agents Chemother. 2014;58:859–864.

33. Chua H, Feng S, Lewnard JA, et al. The use of test-negative controls to monitor vaccine effectiveness. *Epidemiology*. 2020;31:43–64.
34. Foppa IM, Haber M, Ferdinands JM, et al. The case test-negative design for studies of the effectiveness of influenza vaccine. *Vaccine*. 2013;31(30):3104–3109.
35. Lewnard JA, Tedijanto C, Cowling BJ, et al. Measurement of Vaccine Direct Effects under the Test-Negative Design. *Am J Epidemiol*. 2018;187:2686–2697.
36. Mulholland K, Hilton S, Adegbola R, et al. Randomised trial of *Haemophilus influenzae* type-b tetanus protein conjugate for prevention of pneumonia and meningitis in Gambian infants. *Lancet*. 1997;349:1191–1197.
37. Gessner BD, Sutanto A, Linehan M, et al. Incidences of vaccine-preventable *Haemophilus influenzae* type b pneumonia and meningitis in Indonesian children: Hamlet-randomised vaccine-probe trial. *Lancet*. 2005;365:43–52.
38. Halloran ME, Struchiner CJ. Causal inference in infectious diseases. *Epidemiology*. 1995;6(2):142–151.
39. Westreich D, Hudgens MG. Beware the Test-Negative Design. *Am J Epidemiol*. 2016;184(5):354–356.

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
$W_{i,k}$	Observation from the k th confounding variable for individual i
D_i	Indicator that a pathogen is detected ($D_i = 1$) or not detected ($D_i = 0$), for individual i

Table 2: Parameters of the model.

Parameter	Definition	Range
π	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 to 1
θ_P	Prevalence ratio of the pathogen of interest among vaccinated versus unvaccinated controls individuals not experiencing disease symptoms potentially related to this pathogen	0 to 1, for a protective vaccine
ρ	Per-acquisition risk for progression of the pathogen of interest to disease among unvaccinated individuals	0 to 1
θ_S	Risk ratio for progression of the pathogen of interest to disease, given acquisition among vaccinated versus unvaccinated individuals	0 to 1, for a protective vaccine
ω	Risk (measured over the same period as ρ) of disease due to all causes other than the pathogen of interest	0 to 1
v	Proportion of individuals in the population who are vaccinated (equal to the expected prevalence among controls)	0 to 1
β_j	Log-transformed adjusted odds ratio conveying the association of the j^{th} variable with case status	Unbounded
δ_k	Log-transformed adjusted odds ratio conveying the association of the k^{th} confounder with case status	Unbounded

λ	Rate of acquisition of the pathogen of interest among unvaccinated individuals not currently shedding or carrying this pathogen	Greater than 0
γ	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 to 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_ω	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 to 1, for a protective vaccine
α_0	Test sensitivity for detection of the pathogen of interest in an individual experiencing shedding or carriage, without symptoms attributable to the pathogen	0 to 1

α_1 Test sensitivity for detection of the pathogen of interest in an individual experiencing symptoms attributable to the pathogen

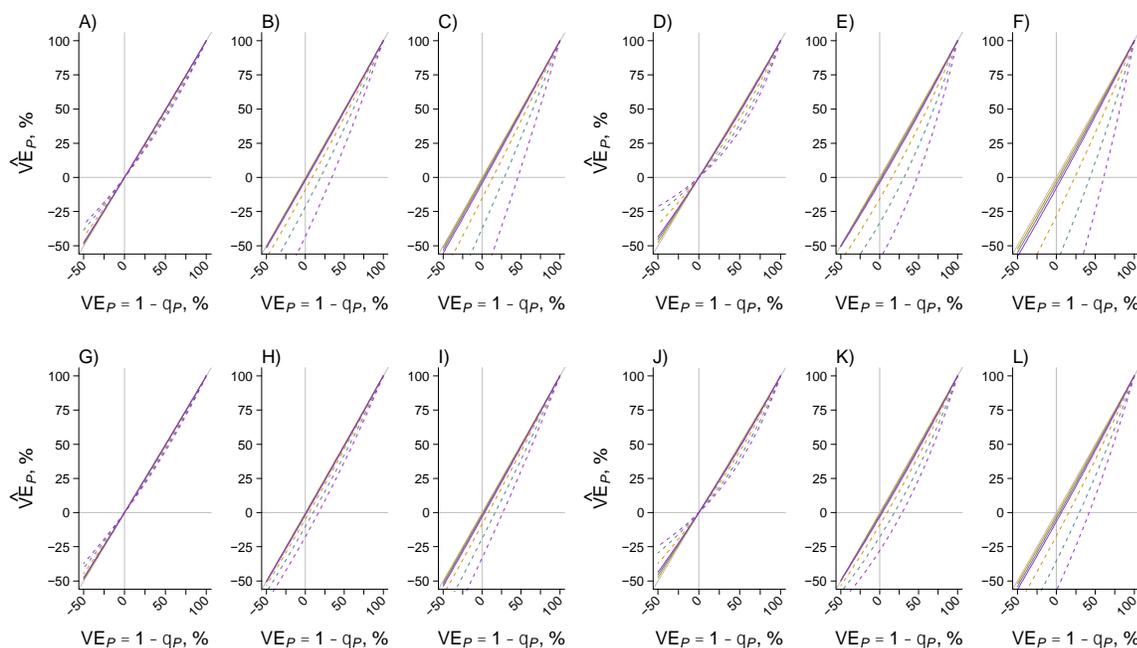


Figure 1: Effect of imperfect test sensitivity on estimates of vaccine-conferred protection against progression of the pathogen to disease. Using the expressions for OR_p^0 and OR_p^1 in Eq. 26, we plot the estimated effect of vaccination on risk of progression of the pathogen of interest, given acquisition ($\widehat{VE}_p = 1 - (OR_p^1 - 1)/(OR_p^0 - 1)$). The x-axis indicates the true vaccine effect against progression; departures from the 1:1 diagonal (grey) line indicate bias. Gold, blue, and violet lines correspond to estimates assuming 90%, 80%, and 70% test sensitivity, respectively, for detecting the pathogen of interest among individuals experiencing disease due to this pathogen (gold, $\alpha_1 = 0.9$; blue, $\alpha_1 = 0.8$; violet, $\alpha_1 = 0.7$). Solid and dashed lines correspond to estimates assuming 10% and 50% prevalence, respectively, of the pathogen of interest among unvaccinated individuals without symptoms (solid, $\pi = 0.1$; dashed, $\pi = 0.5$). The top row

(panels A-F) considers a diagnostic test with 100% sensitivity for detection of the pathogen when it is not a cause of symptoms ($\alpha_0 = 1$), while the bottom row (panels G-L) considers a test with 70% sensitivity for detecting the pathogen when it is not a cause of symptoms ($\alpha_0 = 0.7$).

Within each column, the left three panels (panels A-C and panels G-I) assume acquisition of the pathogen doubles individuals' risk of disease ($\rho/\omega = 1$), while the right three panels (panels D-F and panels J-L) assume acquisition of the pathogen results in a four-fold increase in individuals' risk of disease ($\rho/\omega = 3$). Each grouping of three panels presents estimates corresponding to increasing levels of vaccine-conferred protection against shedding or colonization (VE_S). Left panels (panels A, D, G, J) assume 0% protection ($\theta_S = 1$); center panels (panels B, E, H, K) assume 25% protection ($\theta_S = 0.75$); and right panels (C, F, I, L) assume 50% protection ($\theta_S = 0.5$). Grey lines plotted at $VE_P = 0\%$ and $\widehat{VE}_P = 0\%$ indicate where the true value and estimated value, respectively, cross between indicating increased and decreased risk of disease, given acquisition, as a consequence of vaccination.

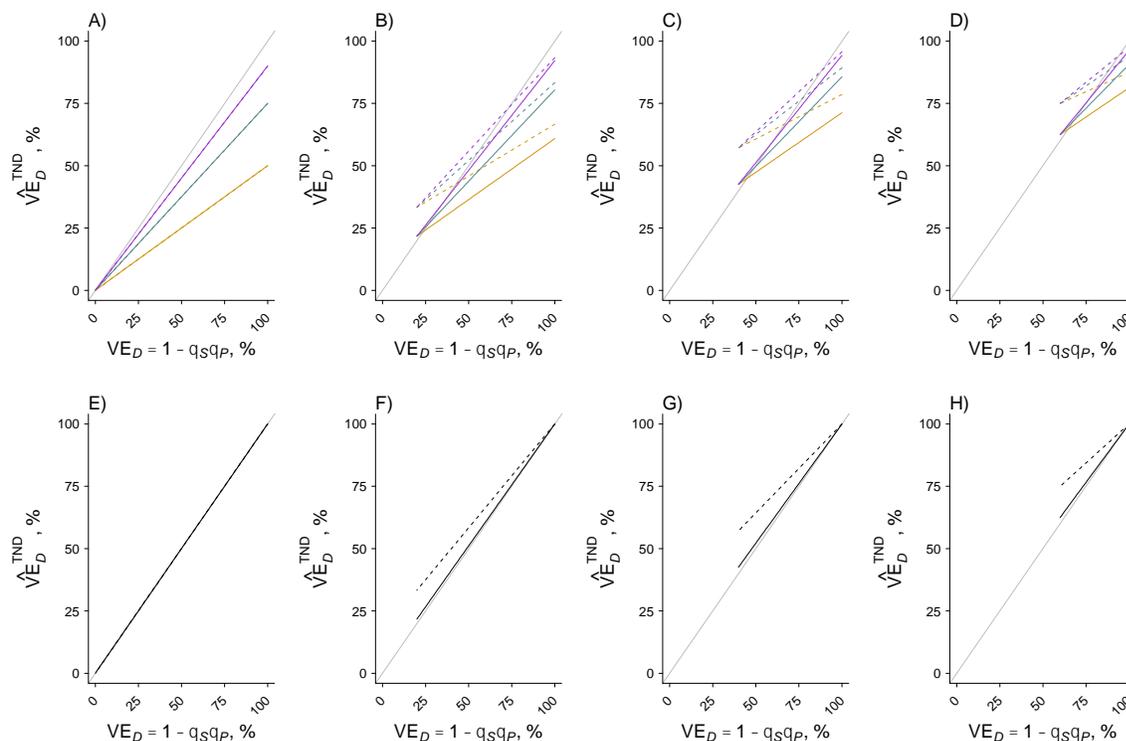


Figure 2: Bias in estimated protection against disease under the test-negative design. We plot the estimated effect of vaccination on risk of disease attributable to the pathogen of interest generated under the test-negative design (\widehat{VE}_D^{TND}), indicating the true vaccine direct effect ($VE_D = 1 - \theta_S \theta_P$) 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 (solid, $\pi = 0.1$; dashed, $\pi = 0.5$). In the top row (panels A-D), gold, blue, and violet lines correspond to estimates assuming acquisition of the pathogen results in a two-fold, four-fold, and ten-fold increase in individuals' risk of disease (gold, $\rho/\omega = 1$; blue, $\rho/\omega = 3$; violet, $\rho/\omega = 9$). Organization of panels from left to right corresponds to estimates with increasing vaccine effectiveness against shedding or carriage of the pathogen of interest (VE_S). The left column (panels A and E) assumes 0% protection ($\theta_S = 1$); the second column (panels B and F) assumes 20% protection ($\theta_S =$

0.75); the third column (panels C and G) assumes 40% protection ($\theta_S = 0.6$); and the right column (panels D and H) assumes 60% protection ($\theta_S = 0.4$).

FOOTNOTES PAGE

Abbreviations:

OR: Odds ratio

TND: Test-negative design

VE: Vaccine effectiveness