

Epidemiology Publish Ahead of Print

DOI: 10.1097/EDE.0000000000001366

Theoretical framework for retrospective studies of the effectiveness of SARS-CoV-2 vaccines

Joseph A. Lewnard,^{1,2,3,*} Manish M. Patel,⁴ Nicholas P. Jewell,^{5,6} Jennifer R. Verani,⁴ Miwako Kobayashi,⁴ Mark W. Tenforde,⁴ Natalie E. Dean,⁷ Benjamin J. Cowling,⁸ Benjamin A. Lopman⁹

1. Division of Epidemiology, School of Public Health, University of California, Berkeley, Berkeley, California, United States
2. Division of Infectious Diseases & Vaccinology, School of Public Health, University of California, Berkeley, Berkeley, California, United States
3. Center for Computational Biology, College of Engineering, University of California, Berkeley, Berkeley, California, United States
4. COVID-19 Response Team, Centers for Disease Control and Prevention, Atlanta, Georgia, USA
5. Department of Medical Statistics, London School of Hygiene and Tropical Medicine, London, United Kingdom
6. Division of Biostatistics, School of Public Health, University of California, Berkeley, Berkeley, California, United States
7. Department of Biostatistics, University of Florida, Gainesville, Florida, United States
8. WHO Collaborating Centre for Infectious Disease Epidemiology and Control, School of Public Health, University of Hong Kong, Hong Kong Special Administrative Region, China

9. Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, Georgia, United States

* Corresponding author: Joseph A. Lewnard 2121 Berkeley Way Room 5410 Berkeley, CA 94720 jlewnard@berkeley.edu +1 (847) 507-6115

Running head: Study designs for SARS-CoV-2 vaccine effectiveness

Disclaimer: The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

Conflicts of interest: JAL has received grants and consulting fees from Pfizer, Inc. unrelated to this research.

Sources of funding: This work was supported by grants R01-AI14812701 from the National Institute for Allergy and Infectious Diseases to NPJ and JAL, and R01-AI139761 from the National Institute for Allergy and Infectious Diseases to NED.

Data and code availability: Code for replication is available from the corresponding author via <https://github.com/joelewnard/covidTND>.

ABSTRACT

Observational studies of the effectiveness of vaccines to prevent COVID-19 are needed to inform real-world use. Such studies are now underway amid the ongoing rollout of SARS-CoV-2 vaccines globally. While traditional case-control and test-negative design studies feature prominently among strategies used to assess vaccine effectiveness, such studies may encounter important threats to validity. Here we review the theoretical basis for estimation of vaccine direct effects under traditional case-control and test-negative design frameworks, addressing specific natural history parameters of SARS-CoV-2 infection and COVID-19 relevant to these designs. Bias may be introduced by misclassification of cases and controls, particularly when clinical case criteria include common, non-specific indicators of COVID-19. When using diagnostic assays with high analytical sensitivity for SARS-CoV-2 detection, individuals testing positive may be counted as cases even if their symptoms are due to other causes. The traditional case-control design may be particularly prone to confounding due to associations of vaccination with healthcare-seeking behavior or risk of infection. The test-negative design reduces but may not eliminate this confounding, for instance if individuals who receive vaccination seek care or testing for less-severe illness. These circumstances indicate the two study designs cannot be applied naively to datasets gathered through public health surveillance or administrative sources. We suggest practical strategies to reduce bias in vaccine effectiveness estimates at the study design and analysis stages.

BACKGROUND

Non-randomized studies undertaken after vaccine authorization or licensure provide crucial information about real-world vaccine effectiveness (VE). Such studies may also address questions not answered by clinical trials, such as the duration of protection, the effectiveness of alternative dosing schedules, and protection against emerging variants or within various population subgroups.^{1,2} While prospective cohort studies provide an opportunity to compare outcomes among vaccinated and unvaccinated individuals, such studies may be logistically prohibitive due to their large size and the complexity of longitudinal follow-up. Thus, retrospective studies comparing prior vaccination among individuals with known clinical outcomes of disease or no disease (i.e. case–control VE studies) provide an efficient alternative.³ Phase III randomized controlled trials (RCTs) have established efficacy of multiple vaccines against COVID-19,^{4–6} with others in progress. Observational VE studies have been prioritized by public health^{7,8} and regulatory⁹ authorities to inform real-world use. Traditional case–control and test-negative design studies are the primary retrospective frameworks for assessing VE.¹⁰ However, these and other observational designs may differ with respect to measures of VE and risks of bias.² Here we lay out the theoretical basis for use of the traditional case–control and test-negative design in assessments of vaccine direct effects against SARS-CoV-2 infection and COVID-19. We highlight key assumptions and potential biases underlying VE estimates from the traditional case–control and test-negative design studies, and provide practical recommendations to support researchers designing, analyzing, and interpreting these studies.

OVERVIEW OF THE DESIGNS

The traditional case–control and test-negative design derive VE estimates by comparing the odds of prior vaccination among individuals who, at the time of enrollment, are cases (experiencing the clinical endpoint against which VE is to be measured) or non-cases (controls). Cases are typically individuals who meet clinical criteria such as the presence of predefined symptoms together with laboratory-confirmed detection of a vaccine-targeted infectious agent.¹¹ Under the traditional case–control, controls are selected from a pool of individuals who are not known to be experiencing the clinical endpoint of interest at the time of enrollment, but are members of the same population from which cases are identified. Such individuals may be community controls, selected from among asymptomatic individuals in the community; registry controls, selected from population-based registries; or healthcare controls, selected among patients experiencing an alternate disease unrelated to the pathogen and vaccine of interest.¹²

The test-negative design is an alternative to the traditional case–control study that has gained popularity for studies of vaccines against influenza, rotavirus, and other pathogens.¹³ In the test-negative design, both cases and controls are selected from among individuals who receive diagnostic tests for a pathogen of interest; conventionally, individuals are tested because they experience a clinical syndrome which may be caused by the vaccine-targeted pathogen or other agents. Cases are those testing positive, and controls are those testing negative, placing a heavy reliance on diagnostic accuracy.^{2,14,15} Since the test-negative does not require active selection of controls from the community or other settings, it may offer logistical advantages over the traditional case–control design.

Below we lay out the general framework for estimation of vaccine direct effects in traditional case–control and test-negative design studies. We will introduce relevant notation and assumptions and then extend this framework to address observations that may be expected in studies of SARS-CoV-2.

IDENTIFICATION OF VACCINE EFFECTIVENESS UNDER TRADITIONAL CASE–CONTROL AND TEST-NEGATIVE DESIGNS

Consider the instantaneous risk (or prevalence) of a clinical condition of interest to be p (in an unvaccinated population), a value between 0-1 (**Table 1**). Consider that vaccination reduces individuals’ risk of the condition by a factor $VE = (1 - \theta)$, such that the relative risk of the condition for vaccinated (versus unvaccinated) individuals is θ ; we define $1 - \theta$ as the vaccine “direct effect” resulting from reduction in biologic susceptibility to the condition of interest due to vaccination.

As measured, VE is a population-average treatment effect; at two extremes, individuals may experience uniform reductions in risk by a factor of $(1 - \theta)$, or a proportion $(1 - \theta)$ of recipients may experience perfect protection while others experience none. Unless otherwise noted, these differing mechanisms of vaccine action (elsewhere termed “leaky” and “all-or-nothing” protection¹⁰, respectively) do not result in practical differences for measurement of VE under the scenarios we consider.

Defining v as the proportion of the population receiving vaccine, the probabilities of individuals falling into various case and control categories, given their vaccination status, is laid out in the two-by-two tables in **Table 2**. Under the traditional case–control design,

$$\widehat{VE}_{TCC} = 1 - \frac{v\theta p \times (1 - v)(1 - p)}{v(1 - \theta p) \times (1 - v)p} = 1 - \frac{\theta(1 - p)}{(1 - \theta p)}$$

which, for rare outcomes ($p \approx 0$), yields an unbiased estimate $\widehat{VE}_{TCC} \approx 1 - \theta = VE$. Alternative design approaches such as incidence density matching of cases and controls¹⁶ or inclusive sampling¹⁷ may reduce reliance on the rare outcome assumption; here for simplicity we focus on more general frameworks.

Under the test-negative design, controls instead experience an outcome that is unaffected by vaccination against the pathogen of interest.^{14,15,18} Defining the risk of this outcome as k , regardless of vaccination (**Table 2**), and assuming a perfect diagnostic test,

$$\widehat{VE}_{TND} = 1 - \frac{v\theta p \times (1 - v)k}{vk \times (1 - v)p} = 1 - \theta = VE,$$

which does not require the rare-disease assumption.

CLASSES OF VACCINE DIRECT EFFECTS

Vaccine-conferred protection may impact multiple aspects of the natural history of infectious disease agents such as SARS-CoV-2, including but not limited to prevention of virus acquisition, reduced viral replication in the upper respiratory mucosa, earlier clearance of infection, and prevention of mild or severe symptoms. The differential effects of SARS-CoV-2 vaccines against infection and disease endpoints remain imperfectly understood¹⁹ and may vary by vaccine product and across virus variants.

Typical estimands of traditional case-control and test-negative design studies are VE against symptomatic disease of any severity, or VE against severe disease. These are also typical target estimands in individually randomized vaccine efficacy trials.²⁰ Define θ_S as the relative risk of infection for a vaccinated versus unvaccinated individual, resulting from reduced susceptibility to acquisition of SARS-CoV-2 or accelerated clearance of the pathogen.²⁰ Define θ_P as the relative risk of COVID-19 (or COVID-19 meeting a particular severity threshold, e.g. hospitalization, intensive care unit admission, or mechanical ventilation) for an infected

vaccinated individual versus an infected unvaccinated individual. Protection resulting from prevention of both infection and progression is $VE_D = 1 - \theta_S \theta_P$; we assume VE_D is the desired estimand and describe biases of \widehat{VE}_{TCC} and \widehat{VE}_{TND} relative to VE_D . Specialized designs would be required to isolate θ_S or θ_P .

OUTCOME MISCLASSIFICATION

Etiologic detection

Clinical outcomes of interest to researchers may range from laboratory confirmation of infection regardless of symptoms^{21–23} to laboratory-confirmed mild, moderate, or severe disease manifestations.²⁰ A positive test result may not specifically indicate the etiology of current symptoms, as SARS-CoV-2 genetic material can be shed for prolonged periods after infection.²⁴ Moreover, many acquisitions of SARS-CoV-2 may never cause symptoms.²⁵

Prolonged viral shedding and the presence of asymptomatic infection may introduce misclassification in both traditional case–control and test-negative design studies. For instance, traditional case–control studies enrolling controls selected from the community, from registries, or among patients seeking care for other (e.g., non-respiratory) diseases, may suffer from misclassification if a proportion of these controls are in fact infected with SARS-CoV-2, but not tested. Inclusion of infected individuals in the control group will diminish the apparent effects of vaccination, provided vaccine prevents infection. Likewise, in both traditional case–control and test-negative design studies, individuals who are shedding SARS-CoV-2 genetic material but experiencing illness due to another cause (e.g., another infection or a non-infectious process with overlapping symptoms such as chronic obstructive pulmonary disease) may be misclassified as cases. Misclassification of asymptotically infected persons as controls, and of infected

symptomatic persons as cases when their illness is not caused by SARS-CoV-2, can be expected to increase with higher prevalence of infection in the community.

To formalize resulting biases, consider separately the risks of SARS-CoV-2 acquisition and disease progression. Defining a as the instantaneous risk (prevalence) of being infected with SARS-CoV-2, at any time, and d as the instantaneous risk of experiencing symptoms, given infection, the risk of symptomatic and asymptomatic SARS-CoV-2 infections are ad and $a(1 - d)$, respectively. We extend our contingency tables to accommodate test-positive symptomatic cases, asymptomatic controls, and test-negative symptomatic controls in **Table 3**; for simplicity we consider binary disease etiology as SARS-CoV-2 or other factors. Here, the traditional case-control yields

$$\widehat{VE}_{TCC} = 1 - \frac{\theta_S[\theta_P d + (1 - \theta_P d)k](1 - ad)}{(1 - \theta_S \theta_P ad)[d + (1 - d)k]},$$

which remains subject to bias owing to the risk of symptoms due to causes other than SARS-CoV-2 (k) and the prevalence of SARS-CoV-2 infection (a). However, these biases may be ignorable under several conditions. If prevalence of other conditions is negligible relative to the risk of COVID-19 for an infected individual ($k \ll d$, if endpoints are chosen with high specificity for COVID-19), the traditional case-control design yields

$$\widehat{VE}_{TCC} \approx 1 - \frac{\theta_S \theta_P (1 - ad)}{1 - \theta_S \theta_P ad}.$$

Such a design may under-estimate VE_D if prevalence of SARS-CoV-2 infection (a) is high owing to inclusion of infected asymptomatic individuals in the control group. If prevalence of SARS-CoV-2 infection in the community is low or negligible ($a \approx 0$), $\widehat{VE}_{TCC} \approx 1 - \theta_S \theta_P$.

Comparing symptomatic individuals who test positive and negative for infection, the test-negative design yields

$$\widehat{VE}_{TND} = 1 - \frac{\theta_S[\theta_P d + (1 - \theta_P d)k](1 - a)}{(1 - \theta_S a)[d + (1 - d)k]}.$$

The same conditions by which bias can be reduced under the case-control design reduce bias under the test-negative design. With $k \ll d$,

$$\widehat{VE}_{TND} \approx 1 - \frac{\theta_S \theta_P (1 - a)}{1 - \theta_S a},$$

and with $a \approx 0$, $\widehat{VE}_{TND} \approx 1 - \theta_S \theta_P$.

We illustrate the quantitative extent of bias under the traditional case-control and test-negative design in **eFigure 1**; <http://links.lww.com/EDE/B808> and **eFigure 2**;

<http://links.lww.com/EDE/B808>. Under most realistic situations, misclassification bias is low or negligible for both designs. While relatively insensitive to changes in prevalence of infection within ranges that have been reported for SARS-CoV-2, bias increases if the association of clinical case-defining symptoms with infection is weak (such that $d \gg k$ is false), potentially leading to substantial under-estimation of VE_D .

Choice of diagnostic assays

Multiple assays are available for SARS-CoV-2 detection, and use of assays with differing performance characteristics may influence the extent of biases identified above. Nucleic acid amplification assays such as polymerase chain reaction (PCR), for instance, have higher analytic sensitivity (i.e., likelihood of positive test for sample containing viral material at the minimum detectable concentration for which assay is designed), potentially leading to detection of SARS-CoV-2 genetic material among individuals with low levels of virus-shedding. Antigen detection tests, with lower analytic sensitivity, may in contrast yield negative results for individuals with

low virus-shedding that remains detectable by molecular testing.²⁶ In comparison to mild illness, both assays may have lower sensitivity for severe disease which can be delayed in presentation, occurring often during the second week after infection onset when viral shedding is lower.⁴² Define σ_P and σ_S as the sensitivity of an assay for detecting SARS-CoV-2 among individuals experiencing COVID-19 and those not experiencing symptoms attributable to SARS-CoV-2, respectively. We limit our consideration to differences in test sensitivity, considering specificity to encompass the related issue of non-etiological SARS-CoV-2 detections; under field conditions, antigen and molecular tests have been found to have similar analytical specificity.^{26,27} The limiting case of an assay which perfectly rules out SARS-CoV-2 detections that are not causing symptoms ($\sigma_S = 0$) yields the contingency table presented in **Table 4** for case and control groups (**eTable 1**; <http://links.lww.com/EDE/B808> extends this for non-zero values of σ_S).

With $\sigma_S = 0$, the traditional case-control yields

$$\widehat{VE}_{TCC} = 1 - \frac{\theta_S \theta_P (1 - ad)}{1 - \theta_S \theta_P ad},$$

removing the reliance on the assumption $k \ll d$, and allowing $\widehat{VE}_{TCC} \approx 1 - \theta_S \theta_P = VE_D$ if overall prevalence of infection is low ($a \approx 0$). Under the test-negative design,

$$\widehat{VE}_{TND} = 1 - \frac{\theta_S \theta_P [k + (1 - k)ad(1 - \sigma_P)]}{k + (1 - k)\theta_S \theta_P ad(1 - \sigma_P)}.$$

Here, we obtain an unbiased estimate $\widehat{VE}_{TND} \approx 1 - \theta_S \theta_P = VE_D$ only with high sensitivity of the assay for individuals experiencing COVID-19 ($\sigma_P \approx 1$). Lower values of σ_P result in greater bias (**eFigure 3**; <http://links.lww.com/EDE/B808>, **eFigure 4**; <http://links.lww.com/EDE/B808>).

Nonetheless, under a scenario where σ_P is high and asymptomatic SARS-CoV-2 detections are minimal ($\sigma_S \approx 0$; **eFigure 3**; <http://links.lww.com/EDE/B808>, **eFigure 4**;

<http://links.lww.com/EDE/B808>), the extent of bias remains lower than what arises with perfect

detection of SARS-CoV-2 in both asymptomatic infection and COVID-19 (as plotted in **eFigure 1**; <http://links.lww.com/EDE/B808>, **eFigure 2**; <http://links.lww.com/EDE/B808>), particularly for vaccines that primarily prevent infection (**eFigure 3**; <http://links.lww.com/EDE/B808>, **eFigure 4**; <http://links.lww.com/EDE/B808>).

We note that no available assay enables perfect differentiation of COVID-19 from non-etiological detections of SARS-CoV-2; statistical corrections for imperfect test sensitivity and specificity have been proposed to adjust final VE estimates.²⁸ Further complicating the selection of assays, available data suggest similar viral loads and/or nucleic acid abundance among individuals with or without symptoms at similar stages of infection.^{29,30} However, onset of COVID-19 upper respiratory tract symptoms typically coincides with peak viral shedding. Thus, individuals seeking care for new-onset COVID-19 may generally have higher viral load than individuals experiencing symptoms owing to other causes, who may be at later stages of SARS-CoV-2 shedding or molecular detection.³¹ This may drive an artefactual association of higher SARS-CoV-2 shedding with COVID-19.

HEALTHCARE-SEEKING BEHAVIOR

Differential healthcare-seeking behaviors of individuals who can access and receive vaccines, versus others, may be another source of bias. The test-negative design aims to limit such bias by restricting enrollment to individuals who seek care and receive diagnostic testing for the same or similar clinical indications. In traditional case–control studies, there may be uncertainty about whether controls were truly disease-free, as some may not have sought care given symptoms. This may be particularly true if studies select controls from registries without direct outreach.

We may quantify the resulting bias by considering the number of controls who are misclassified when the probability of seeking care, given symptoms, is π_V among the vaccinated and π_U among the unvaccinated. Defining eligible controls as individuals who do not become ascertained as cases (**Table 5**):

$$\widehat{VE}_{TCC} = 1 - \frac{\theta_S[\theta_P d + (1 - \theta_P d)k][1 - (k + (1 - k)ad)\pi_U]}{[d + (1 - d)k][1 - (k + (1 - k)\theta_S\theta_P ad)\pi_V]} \times \frac{\pi_V}{\pi_U}.$$

The resulting bias generally leads to under-estimation of true protection (VE_D) when the probability of seeking treatment is higher among vaccinated than unvaccinated persons, given the same status of being infected or symptomatic, owing to the disproportionately higher likelihood for vaccinated individuals to be ascertained as cases (**eFigure 5**;

<http://links.lww.com/EDE/B808>, eFigure 6; <http://links.lww.com/EDE/B808>). Studies aiming to mitigate this bias by enrolling health facility controls may still encounter difficulty. If vaccination is associated with higher rates of healthcare seeking, prevalence of vaccination among controls seeking care for less-severe conditions may be inflated, leading to overestimation of VE_D .

Applying the same formulation under the test-negative design, the terms π_V and π_U cancel out (**Table 5**). Nonetheless, in practical contexts, test-negative studies may enroll individuals across a spectrum of clinical severity. This may cause bias to persist if the probability of treatment-seeking differs for vaccinated and unvaccinated individuals with relatively less severe or more severe illness.¹⁸ These considerations motivate approaches such as matching of cases and test-negative controls (or cases and alternative-disease controls under the traditional case-control) on clinical severity.

Prevalence of other causes of illness leading to care-seeking (e.g., among alternative-disease controls under the traditional case–control and among test-negative controls under the test-negative design) may vary considerably over time. If risk of disease due to these other causes is also associated with likelihood of vaccination, further bias may arise. Even under normal circumstances, disease etiology among controls can change within a season or from year to year. This variation is likely to be extreme over the time horizon that SARS-CoV-2 vaccines are rolled out, as non-pharmaceutical interventions targeting SARS-CoV-2 have affected circulation of other pathogens.^{32,33} As non-pharmaceutical interventions are rolled back, other causes of healthcare-seeking may surge. Since many infections causing respiratory symptoms (e.g., influenza, pneumococcus) are vaccine preventable, controls with these infections may be less likely to have sought vaccine. Accordingly, the ratio π_V/π_U may change over time.

A CAUTIONARY NOTE

Testing for SARS-CoV-2 is commonly undertaken for both clinical diagnostic and public health purposes. Thus, data on SARS-CoV-2 test results may encompass individuals tested because of clinical symptoms as well as those tested due to contact with a known or suspected COVID-19 case, for purposes of screening before work or medical procedures, or for other indications. Stratifying analyses to address fairly homogeneous populations, or to individuals who receive SARS-CoV-2 testing with the same frequency or for the same indications, may be of considerable importance.

Variation in the likelihood of seeking testing is expected to be greatest for cases of mild illness or among persons without symptoms. There may be substantial associations to test-seeking with individuals' healthcare utilization, including the propensity to be vaccinated. In the most extreme case, populations targeted for vaccination may also be populations targeted for frequent testing,

such as healthcare providers, residents of long-term care facilities, and essential workers. Infections in these groups will be more likely to be detected relative to infections within groups without similar access to testing, who may also be less likely to be vaccinated. Here $\pi_V \gg \pi_U$, potentially leading to substantial under-estimation of VE. Similar bias may persist when vaccines are made more widely available if individuals with the greatest propensity to seek healthcare, given the same clinical presentation, including SARS-CoV-2 testing, are also more likely to be vaccinated.

Thus, the test-negative design cannot be applied naively to SARS-CoV-2 testing datasets gathered through public health surveillance or administrative data sources. To adequately correct for potential differences in vaccinated and unvaccinated individuals' likelihood of being tested, designs or analyses should compare individuals being tested for the same indications (e.g., fulfilling particular clinical criteria, known high-risk exposure to SARS-CoV-2, or routine screening), for instance by matching or stratifying on vaccine eligibility criteria. Bias may be further reduced in studies considering severe disease endpoints for which both vaccinated and unvaccinated individuals would be unlikely to defer treatment, regardless of their propensity to seek healthcare for less severe illness.

ASSOCIATION OF VACCINE UPTAKE WITH DIFFERENTIAL RISK OF INFECTION

Confounding in prevalence of infection and immunity

Limited vaccine supplies are being targeted initially at groups most likely to be exposed to SARS-CoV-2 as well as those at greatest risk of severe COVID-19, if infected. Even when vaccines are more widely available, uptake of vaccination may be highest among individuals who adhered most strongly to non-pharmaceutical interventions. Accordingly, exposure to

infection and risk of progressing to severe disease will not be equal between vaccinated and unvaccinated groups. This differential extends not only to rates of new infection, but also to the prevalence of prior natural infection and immunity. As available evidence suggests naturally acquired immunity is strongly (~90%) protective against re-infection, there may be substantial differences in risk among vaccinated and unvaccinated individuals owing to factors other than vaccine-derived protection.^{34,35}

We extend the contingency tables laid out previously (**Table 6**) to incorporate β , the relative risk of SARS-CoV-2 infection among vaccinated (vs. unvaccinated) individuals due to factors other than vaccination, as well as ω_V and ω_U , the proportions of the vaccinated and unvaccinated populations, respectively, that remain uninfected. For simplicity, we assume near-total protection against reinfection conferred by naturally acquired immunity. Under the two designs,

$$\widehat{VE}_{TCC} = 1 - \frac{\theta_S \beta [\theta_P d + (1 - \theta_P d)k] \omega_V (1 - \omega_U a d)}{[d + (1 - d)k] \omega_U (1 - \theta_S \theta_P \beta a d \omega_V)}$$

and

$$\widehat{VE}_{TND} = 1 - \frac{\theta_S \beta [\theta_P d + (1 - \theta_P d)k] \omega_V (1 - a \omega_U)}{[d + (1 - d)k] \omega_U (1 - \theta_S \beta a \omega_V)},$$

which, with low prevalence of infection ($a \approx 0$) and low risk of disease due to other causes ($k \ll d$), yield

$$\widehat{VE}_{TCC} = \widehat{VE}_{TND} \approx 1 - \theta_S \theta_P \times \beta \frac{\omega_V}{\omega_U}.$$

Here, differential risk of SARS-CoV-2 infection and prevalence of immunity among vaccinated and unvaccinated individuals lead to counteracting biases. With higher risk among individuals who receive vaccination ($\beta > 1$), we may expect higher prevalence of naturally acquired infection in this group and $\omega_V < \omega_U$; bias is offset fully if $\beta = \omega_U / \omega_V$.

We plot conditions under which bias is driven predominantly by differential exposure to infection or by differential naturally acquired immunity in **Figure 1**. Considering the context of VE studies undertaken shortly after vaccine implementation, we may assume individuals' time at risk is contributed primarily in the period before vaccination; with exponentially distributed infection times, and a cumulative force of infection c exerted over the period at risk, we may define the remaining uninfected populations $\omega_U = \exp(-c)$ and $\omega_V = \exp(-\beta c)$. Bias is fully offset with $c = \log(\beta) / (\beta - 1)$, yielding $\omega_U = \beta^{\frac{1}{1-\beta}}$ (**Figure 1**). Thus, the prevalence of pre-existing naturally acquired immunity needed to offset bias resulting from differential exposure is high, at 33% and 56% for $\beta = 5$ and $\beta = 1.5$, respectively.

Detailed risk factor data are likely necessary to address confounding associated with differential risk of infection among individuals receiving and those not receiving vaccination. While certain occupations (e.g. healthcare or other essential workers) may be associated with heightened risk of exposure relative to the general public, substantial variation in risk may exist within these categories owing to differences in the specific tasks conducted and availability of as well as adherence to infection control measures. Individuals may modify aspects of their behavior following vaccine receipt, further altering differences in risk of infection among vaccine recipients and non-recipients. Collection of questionnaire items addressing risk behaviors and exposures is thus of importance to mitigate confounding.

Temporal considerations

Tiered rollout of vaccination while disease incidence changes concurrently will cause the bias we have described above to vary over time. Thus, in addition to differential exposure or susceptibility, cohorts with earlier or later access to and uptake of vaccines may be exposed to differing risk of infection at equivalent time points after vaccination. Matching, stratification, or

statistical adjustment may be necessary to correct for temporal confounding of this nature, and to avoid spurious signals of time-varying VE. While we have focused here on studies undertaken shortly after vaccine rollout, prior work has considered differential accrual of natural immunity among vaccinated and unvaccinated individuals over longer periods of time.^{3,18,36,37} For vaccines conferring “leaky” protection, traditional case–control and test-negative design studies may underestimate true protection owing to higher rates of acquisition of naturally-acquired immunity among unvaccinated individuals. This bias does not arise, however, for vaccines conferring “all-or-nothing” protection.^{3,18}

DISCUSSION

Our analysis provides several practical insights to inform upcoming traditional case–control and test-negative design studies of SARS-CoV-2 vaccines. While we identify conditions and reasonable assumptions under which both designs can supply reliable VE estimates, several threats to validity should be taken into account in the design, analysis, and interpretation of traditional case–control and test-negative design studies. Below we offer considerations for researchers undertaking studies of either design to investigate vaccines against SARS-CoV-2:

Use of severe COVID-19 clinical endpoints: Use of severe disease endpoints minimizes risk of misclassification and healthcare-seeking bias. To allow sufficient power, studies may need to enroll patients across multiple healthcare facilities, underscoring the need for standardized clinical criteria and case definitions. Studies addressing VE against common or less-severe clinical endpoints may benefit from using extended analytic methods that have been proposed previously for estimation of VE against diseases with non-specific diagnostic criteria.^{11,38,39} As these analysis methods must generally account for the strength of association between SARS-CoV-2 infection and clinical symptoms, enrollment and testing of control groups without

symptoms related to SARS-CoV-2 infection may be of value for researchers studying VE against such endpoints.

Use of molecular or antigen tests: Studies using SARS-CoV-2 detection assays with the highest analytic sensitivity may encounter the greatest risk of misclassifying symptomatic individuals at the end of their course of infection as cases, including when SARS-CoV-2 is not the cause of symptoms. This could be more problematic when prevalence of SARS-CoV-2 infection is high and outcomes of interest include non-severe or non-specific COVID-19 clinical manifestations, such as in outpatient settings. However, trade-offs associated with risk of bias must be balanced against assay specificity and test performance under varied prevalence and clinical settings. Where molecular assays with high analytical sensitivity are preferred for clinical care provision, researchers may consider analyzing quantitative cycle threshold data or collecting a second specimen for antigen-based testing.

Collecting detailed information on clinical severity and testing indications: Reliance on clinic-based testing under both designs makes it difficult to identify and correct for differences in the likelihood for vaccinated and unvaccinated individuals, given infection and symptoms, to be ascertained as cases. Collection of standardized data on specific clinical symptoms (and their duration) among both cases and controls may reduce the risk of bias related to vaccine-associated differences in healthcare access or healthcare seeking. This strategy further provides an opportunity to identify and exclude, if applicable, registry-based or community controls who were ill but did not seek care. If testing is conducted for multiple pathogens, researchers may also investigate bias by assessing VE against other illnesses not prevented by SARS-CoV-2 vaccination as a “sham” analysis to indicate risk of bias.^{40,41,40,41}

Background immunity: Prevalent naturally acquired immunity may complicate the interpretation of results under both the traditional case–control and TND, potentially contributing to lower risk of infection and disease among individuals with the greatest risk of SARS-CoV-2 exposure. As initial vaccine doses will be prioritized to groups with high infection risk, individuals eligible to receive vaccine at the earliest opportunity may have higher prevalence of pre-existing naturally acquired immunity than those in lower-priority tiers. Studies may mitigate this bias, in part, by comparing individuals within the same prioritization groups, by collecting detailed risk factor data to control for variation in exposure, and through use of serological assays (if available) that differentiate vaccine-derived or naturally induced responses (e.g. to non-vaccine antigens).

REFERENCES

- 1 Hanquet G, Valenciano M, Simondon F, Moren A. Vaccine effects and impact of vaccination programmes in post-licensure studies. *Vaccine* 2013; DOI:10.1016/j.vaccine.2013.07.006.
- 2 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**: 2060–74.
- 3 Smith PG, Rodrigues LC, Fine PEM. Assessment of the protective efficacy of vaccines against common diseases using case–control and cohort studies. *Int J Epidemiol* 1984; **13**: 87–93.
- 4 Polack FP, Thomas SJ, Kitchin N, *et al.* Safety and efficacy of the BNT162b2 mRNA COVID-19 vaccine. *N Engl J Med* 2020; **383**: 2603–15.
- 5 Spector SA, Roupheal N, Creech CB, *et al.* Efficacy and safety of the mRNA-1273 SARS-CoV-2 vaccine. *N Engl J Med* 2020; doi:10.1056/NEJMoa2035389.
- 6 Voysey M, Clemens SAC, Madhi SA, *et al.* Safety and efficacy of the ChAdOx1 nCoV-19 vaccine (AZD1222) against SARS-CoV-2: an interim analysis of four randomised controlled trials in Brazil, South Africa, and the UK. *Lancet* 2020; **397**: 99–111.
- 7 Patel MM, Jackson ML, Ferdinands J. Postlicensure evaluation of COVID-19 vaccines. *JAMA* 2020; DOI:10.1001/jama.2020.19328.
- 8 Hodgson SH, Mansatta K, Mallett G, Harris V, Emary KRW, Pollard AJ. What defines an efficacious COVID-19 vaccine? A review of the challenges assessing the clinical efficacy of vaccines against SARS-CoV-2. *Lancet Infect Dis* 2020; DOI:10.1016/S1473-3099(20)30773-8.

- 9 FDA Center for Biologics Evaluation and Research. Development and licensure of vaccines to prevent COVID-19: Guidance for industry. 2020. Available from: <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/development-and-licensure-vaccines-prevent-covid-19>.
- 10 Halloran ME, Longini IM, Struchiner CJ. Design and interpretation of vaccine field studies. *Epidemiol Rev* 1999; **21**: 73–88.
- 11 Lewnard JA. Uses of pathogen detection data to estimate vaccine direct effects in case–control studies. *J R Soc Interface* 2020; **17**: 20200161.
- 12 Wacholder S, Silverman DT, Mclaughlin JK, Mandel JS. Selection of controls in case–control studies: II. Types of controls. *Am J Epidemiol* 1992; **135**: 1029–41.
- 13 Chua H, Feng S, Lewnard JA, *et al*. The use of test-negative controls to monitor vaccine effectiveness. *Epidemiology* 2020; **31**: 43–64.
- 14 Jackson ML, Nelson JC. The test-negative design for estimating influenza vaccine effectiveness. *Vaccine* 2013; **31**: 2165–8.
- 15 Foppa IM, Haber M, Ferdinands JM, Shay DK. The case test-negative design for studies of the effectiveness of influenza vaccine. *Vaccine* 2013; **31**: 3104–9.
- 16 Greenland S, Thomas DC. On the need for the rare disease assumption in case–control studies. *Am J Epidemiol* 1982; DOI:10.1093/oxfordjournals.aje.a113439.
- 17 Rodrigues LC, Smith PG. Use of the case–control approach in vaccine evaluation: efficacy and adverse effects. *Epidemiol Rev* 1999; **21**: 56–72.
- 18 Lewnard JA, Tedijanto C, Cowling BJ, Lipsitch M. Measurement of vaccine direct effects under the test-negative design. *Am J Epidemiol* 2018; DOI:10.1093/aje/kwy163.

- 19 Lipsitch M, Dean NE. Understanding COVID-19 vaccine efficacy. *Science* 2020; DOI:10.1126/science.abe5938.
- 20 Halloran ME, Struchiner CJ, Longini IM. Study designs for evaluating different efficacy and effectiveness aspects of vaccines. *Am J Epidemiol* 1997; **146**: 789–803.
- 21 Harper DM, Franco EL, Wheeler C, *et al.* Efficacy of a bivalent L1 virus-like particle vaccine in prevention of infection with human papillomavirus types 16 and 18 in young women: a randomised controlled trial. *Lancet* 2004; **364**: 1757–65.
- 22 Alonso PL, Sacarlal J, Aponte JJ, *et al.* Efficacy of the RTS,S/AS02A vaccine against *Plasmodium falciparum* infection and disease in young African children: randomised controlled trial. *Lancet* 2004; **364**: 1411–20.
- 23 Collins AM, Wright AD, Mitsi E, *et al.* First human challenge testing of a pneumococcal vaccine: Double-blind randomized controlled trial. *Am J Respir Crit Care Med* 2015; DOI:10.1164/rccm.201503-0542OC.
- 24 Cevik M, Tate M, Lloyd O, Maraolo AE, Schafers J, Ho A. SARS-CoV-2, SARS-CoV, and MERS-CoV viral load dynamics, duration of viral shedding, and infectiousness: a systematic review and meta-analysis. *Lancet Microbe* 2020; DOI:10.1016/s2666-5247(20)30172-5.
- 25 Buitrago-Garcia D, Egli-Gany D, Counotte MJ, *et al.* Occurrence and transmission potential of asymptomatic and presymptomatic SARSCoV-2 infections: A living systematic review and meta-analysis. *PLoS Med* 2020; DOI:10.1371/journal.pmed.1003346.
- 26 Pilarowski G, Marquez C, Rubio L, *et al.* Field performance and public health response using the BinaxNOW™ Rapid SARS-CoV-2 antigen detection assay during

- community-based testing. *Clin Infect Dis* 2020; DOI:10.1093/cid/ciaa1890.
- 27 Albert E, Torres I, Bueno F, *et al.* Field evaluation of a rapid antigen test (Panbio™ COVID-19 Ag Rapid Test Device) for COVID-19 diagnosis in primary healthcare centres. *Clin Microbiol Infect* 2020; DOI:10.1016/j.cmi.2020.11.004.
- 28 Endo A, Funk S, Kucharski A. Bias correction methods for test-negative designs in the presence of misclassification. *Epidemiol Infect* 2019; DOI:10.1101/19002691.
- 29 Van Vinh Chau N, Lam VT, Dung NT, *et al.* The natural history and transmission potential of asymptomatic severe acute respiratory syndrome coronavirus 2 infection. *Clin Infect Dis* 2020; **71**: 2679–87.
- 30 Long QX, Liu BZ, Deng HJ, *et al.* Antibody responses to SARS-CoV-2 in patients with COVID-19. *Nat Med* 2020; DOI:10.1038/s41591-020-0897-1.
- 31 Zou L, Ruan F, Huang M, *et al.* SARS-CoV-2 viral load in upper respiratory specimens of infected patients. *N Engl J Med* 2020; DOI:10.1056/NEJMc2001737.
- 32 Cowling BJ, Ali ST, Ng TWY, *et al.* Impact assessment of non-pharmaceutical interventions against coronavirus disease 2019 and influenza in Hong Kong: an observational study. *Lancet Public Heal* 2020; DOI:10.1016/S2468-2667(20)30090-6.
- 33 Solomon DA, Sherman AC, Kanjilal S. Influenza in the COVID-19 Era. *JAMA* 2020; DOI:10.1001/jama.2020.14661.
- 34 Marsden BD, Cox S, James T, *et al.* Antibody status and incidence of SARS-CoV-2 infection in health care workers. *N Engl J Med* 2020; DOI:10.1056/NEJMoa2034545.
- 35 Addetia A, Crawford KHD, Dingens A, *et al.* Neutralizing antibodies correlate with protection from SARS-CoV-2 in humans during a fishery vessel outbreak with a high attack rate. *J Clin Microbiol* 2020; **58**: 1–11.

- 36 Farrington CP. The measurement and interpretation of age-specific vaccine efficacy. *Int J Epidemiol* 1992; **21**: 1014–20.
- 37 Lopman BA, Pitzer VE. Waxing Understanding of Waning Immunity. *J Infect Dis* 2018; **217**: 851–3.
- 38 De Smedt T, Merrall E, Macina D, Perez-Vilar S, Andrews N, Bollaerts K. Bias due to differential and non-differential disease- And exposure misclassification in studies of vaccine effectiveness. *PLoS One* 2018. DOI:10.1371/journal.pone.0199180.
- 39 Hasegawa RB, Small DS. Estimating malaria vaccine efficacy in the absence of a gold standard case definition: Mendelian factorial design. *J Am Stat Assoc* 2020; doi:10.1080/01621459.2020.1863222.
- 40 Lipsitch M, Tchetgen Tchetgen E, Cohen T. Negative Controls: A tool for detecting confounding and bias in observational studies. *Epidemiology* 2010. DOI:10.1097/EDE.0b013e3181d61eeb.
- 41 Lewnard JA, Lo NC, Arinaminpathy N, Frost I, Laxminarayan R. Childhood vaccines and antibiotic use in low and middle income countries. *Nature* 2020; **581**: 94–9.
- 42 Lewnard JA, Mora AM, Nkwocha O, *et al.* Prevalence and clinical profile of SARS-CoV-2 infection among farmworkers in Monterey County, California: June–November, 2020. *medRxiv* 2020.
- 43 National Center for Health Statistics. Hospitals, beds, and occupancy rates, by type of ownership and size of hospital: United States, selected years 1975–2015. 2017; Available from: <https://www.cdc.gov/nchs/data/hus/2017/089.pdf>.
- 44 Halpern NA, Pastores SM. Critical care medicine beds, use, occupancy, and costs in the United States: A methodological review. *Crit Care Med* 2015;

DOI:10.1097/CCM.0000000000001227.

- 45 Menachemi N, Yiannoutsos CT, Dixon BE, *et al.* Population point prevalence of SARS-CoV-2 infection based on a statewide random sample — Indiana, April 25–29, 2020. *MMWR* 2020; DOI:10.15585/mmwr.mm6929e1.
- 46 Riley S, Ainslie KEC, Eales O, *et al.* REACT-1 round 6 updated report high prevalence of SARS-CoV-2 swab positivity with reduced rate of growth in England at the start of November 2020. *medRxiv* 2020; doi:10.1101/2020.11.18.20233932.
- 47 Chamie G, Marquez C, Crawford E, *et al.* Community transmission of severe acute respiratory syndrome coronavirus 2 disproportionately affects the Latinx population during shelter-in-place in San Francisco. *Clin Infect Dis* 2020; doi:10.1093/cid/ciaa1234.
- 48 Salje H, Tran Kiem C, Lefrancq N, *et al.* Estimating the burden of SARS-CoV-2 in France. *Science* 2020; DOI:10.1126/science.abc3517.
- 49 Lewnard JA, Liu VX, Jackson ML, *et al.* Incidence, clinical outcomes, and transmission dynamics of severe coronavirus disease 2019 in California and Washington, United States: prospective cohort study. *BMJ* 2020; doi:2020.04.12.20062943.

FIGURE CAPTIONS

Figure 1: Estimated vaccine effectiveness under the traditional case–control design and test-negative design with differential exposure to infection and prevalence of naturally acquired immunity among vaccinated and unvaccinated persons. We illustrate expected estimates of the odds ratio of vaccination given case versus control (or test-negative) status, under the limiting case of $a \approx 0$ and $k \approx 0$, under scenarios where vaccination occurs among individuals with higher (left panel: $\beta = 2$, solid lines; $\beta = 4$, dashed lines) or lower (center panel: $\beta = 0.25$, solid lines; $\beta = 0.5$, dashed lines) risk of exposure to SARS-CoV-2. Colors correspond to differing proportions of the unvaccinated population remaining susceptible to infection ($\omega_U = \exp(-c)$); we define the proportion of the vaccinated population remaining uninfected as $\omega_V = \exp(-\beta c)$ for given values of ω_U . Red diagonal lines across the left and center panels illustrate the true effect. The right panel illustrates threshold values of $\omega_U^* = \beta^{\frac{1}{1-\beta}}$, at which the effects of differential exposure to SARS-CoV-2 and naturally-acquired immunity among vaccinated versus unvaccinated persons cancel out, resulting in a change in the direction of bias.

Table 1: Parameters and their definitions.

Parameter	Definition	Range of values
ρ	Prevalence of a vaccine-preventable clinical condition of interest among unvaccinated individuals for the general setting of traditional case-control and test-negative design studies.	0 to 1
θ	Risk ratio for a clinical outcome of interest among counterfactually vaccinated versus unvaccinated individuals; we define θ_s as the risk ratio of shedding the pathogen (owing to differential acquisition among the vaccinated) and θ_p as the risk ratio of disease progression, given infection. Vaccine effectiveness is defined as $VE = 1 - \theta$ in the general case. The combined extent of protection resulting from prevention of shedding and progression is $VE_D = 1 - \theta_s \theta_p$.	$0 \leq \theta < 1$ corresponds to the scenario of a protective vaccine
ν	Vaccine uptake, or the proportion of the population being studied who have received vaccination.	0 to 1
k	Prevalence of the clinical outcome of interest in the population being studied, attributable to causes other than the vaccine-targeted pathogen; ideally, $k \approx 0$ or $k \ll d$. To illustrate the impact of this parameter, we consider three values corresponding to differing endpoints. We consider $k = 0.184$ for the prevalence of any COVID-19-like symptoms among uninfected persons in a community sample ⁴² . We consider $k = 0.0018$ for the prevalence of hospitalization due to causes other than COVID-19, and $k = 0.00021$ for the prevalence of intensive care utilization due to causes other than COVID-19, based the average number of occupied hospital and intensive care beds per capita in the US. ^{43,44}	0 to 1
a	Prevalence of SARS-CoV-2 shedding among all members of the population being studied. Generally, $a \leq 0.01$ throughout the pandemic, although higher values have been reported in certain high-risk populations. ^{42,45-47} To illustrate the impact of this parameter, we plot scenarios with low ($a = 0.01$) and high ($a = 0.1$) values.	0 to 1
d	Proportion of infected persons experiencing the COVID-19 clinical outcome within the population being studied. To illustrate the impact of this parameter, we consider three values corresponding to differing endpoints. We assume $d = 0.5$ for any symptoms; $d = 0.029$ for hospitalization ⁴⁸ ; and $d = 0.012$ for ICU admission. ⁴⁹	0 to 1
σ	Test sensitivity, or the probability of SARS-CoV-2 detection in an infected patient. We further define σ_s as sensitivity for detecting SARS-CoV-2 when the pathogen is not causing symptoms, and σ_p as sensitivity for detecting SARS-CoV-2 when the pathogen is the cause of symptoms.	0 to 1
π	Probability that individuals who experience the clinical outcome of interest will receive a test. We define π_V and π_U for vaccinated and unvaccinated individuals, respectively, to account for potential differences in healthcare access or healthcare seeking associated with vaccination.	0 to 1
β	Relative risk of SARS-CoV-2 infection among vaccinated (versus unvaccinated) individuals, owing the factors other than vaccination (e.g., differential exposure or susceptibility to infection).	$\beta > 1$ corresponds to a scenario of higher risk among those who will receive vaccination; $0 \leq \beta < 1$ corresponds to a scenario of lower risk among the vaccinated.
ω	Proportion of the population that remains uninfected. We define ω_V and ω_U for vaccinated and unvaccinated individuals, respectively, to account for potential differences in risk of infection within these populations, apart from vaccine-conferred immunity. For the case of exponentially distributed infection times, we consider $\omega_U = \exp(-c)$ and $\omega_V = \exp(-\beta c)$.	0 to 1

Table 2: Two-by-two tables for traditional case-control and test-negative design studies.

Design	Exposure	Outcome	
Traditional case-control	Vaccinated	<u>Case</u> $v\theta p$	<u>Control</u> $v(1-\theta p)$
	Unvaccinated	$(1-v)p$	$(1-v)(1-p)$
Test-negative design	Vaccinated	<u>Test-positive</u> $v\theta p$	<u>Test-negative</u> vk
	Unvaccinated	$(1-v)p$	$(1-v)k$

Contents of the table are probabilities of the various listed outcomes, accounting for vaccinated or unvaccinated status. Parameters: v , proportion of the population vaccinated; p , risk of the outcome of interest; k , risk of the negative-control outcome; θ , relative risk of the outcome of interest, given vaccination (see **Table 1**).

ACCEPTED

Table 3: Contingency tables for traditional case–control and test-negative design studies with outcome misclassification.

Exposure	Outcome		
	Test-positive case (symptomatic)	Community control (asymptomatic)	Test-negative control (symptomatic)
Vaccinated	$v \theta_s a [\theta_p d + (1 - \theta_p d) k]$	$v(1 - k)(1 - \theta_s \theta_p a d)$	$v(1 - \theta_s a) k$
Unvaccinated	$(1 - v) a [d + (1 - d) k]$	$(1 - v)(1 - k)(1 - a d)$	$(1 - v)(1 - a) k$

Contents of the table are probabilities of the various listed outcomes, accounting for vaccinated or unvaccinated status. Parameters: v , proportion of the population vaccinated; a , prevalence of SARS-CoV-2 infection; d , risk of COVID-19 among those with SARS-CoV-2 infection; k , risk of symptoms due to causes other than SARS-CoV-2 infection; θ_s , relative risk of SARS-CoV-2, given vaccination; θ_p , relative risk of COVID-19, among those with SARS-CoV-2 infection, given vaccination (**Table 1**).

ACCEPTED

Table 4: Contingency tables for traditional case–control and test-negative design studies with exclusion of SARS-CoV-2 shedding from COVID-19.

Exposure	Outcome		
	Test-positive case (symptomatic)	Community control (asymptomatic)	Test-negative control (symptomatic)
Vaccinated	$v \theta_s \theta_p a d \sigma_p$	$v(1-k)(1-\theta_s \theta_p a d)$	$v[k + (1-k)\theta_s \theta_p a d(1-\sigma_p)]$
Unvaccinated	$(1-v)a d \sigma_p$	$(1-v)(1-k)(1-a d)$	$(1-v)[k + (1-k)a d(1-\sigma_p)]$

Contents of the table are probabilities of the various listed outcomes, accounting for vaccinated or unvaccinated status. Parameters: v , proportion of the population vaccinated; a , prevalence of SARS-CoV-2 infection; d , risk of COVID-19 among those with SARS-CoV-2 infection; k , risk of symptoms due to causes other than SARS-CoV-2 infection; θ_s , relative risk of SARS-CoV-2, given vaccination; θ_p , relative risk of COVID-19, among those with SARS-CoV-2 infection, given vaccination; σ_p , sensitivity for SARS-CoV-2 detection among individuals experiencing COVID-19 (see **Table 1**). We present a more general contingency table for non-zero values of σ_s (test sensitivity for SARS-CoV-2 infection among those not experiencing COVID-19) in **eTable S1**.

ACCEPTED

Table 5: Contingency table for traditional case–control and test-negative design studies with differential healthcare-seeking among the vaccinated and unvaccinated.

Exposure	Outcome		
	Test-positive case (symptomatic, healthcare sought)	Community control (healthcare not sought)	Test-negative control (symptomatic, healthcare sought)
Vaccinated	$v \theta_s a [\theta_p d + (1 - \theta_p d) k] \pi_v$	$v [1 - (k + \theta_s \theta_p a d (1 - k)) \pi_v]$	$v (1 - \theta_s a) k \pi_v$
Unvaccinated	$(1 - v) a [d + (1 - d) k] \pi_u$	$(1 - v) [1 - (k + a d (1 - k)) \pi_u]$	$(1 - v) (1 - a) k \pi_u$

Contents of the table are probabilities of the various listed outcomes, accounting for vaccinated or unvaccinated status. Parameters: v , proportion of the population vaccinated; a , prevalence of SARS-CoV-2 infection; d , risk of COVID-19 among those with SARS-CoV-2 infection; k , risk of symptoms due to causes other than SARS-CoV-2 infection; θ_s , relative risk of SARS-CoV-2, given vaccination; θ_p , relative risk of COVID-19, among those with SARS-CoV-2 infection, given vaccination; π_v , probability of testing for symptomatic vaccinated individuals; π_u , probability of testing for symptomatic unvaccinated individuals (Table 1).

ACCEPTED

Table 6: Contingency table for traditional case–control and test-negative design studies with differential risk of current and prior infection among the vaccinated and unvaccinated.

Exposure	Outcome		
	<u>Test-positive case (symptomatic)</u>	<u>Community control (asymptomatic)</u>	<u>Test-negative control (symptomatic)</u>
Vaccinated	$v\theta_s\beta a[\theta_p d + (1 - \theta_p d)k]\omega_v$	$v(1 - k)(1 - \theta_s\theta_p\beta ad\omega_v)$	$v(1 - \theta_s\beta a\omega_v)k$
Unvaccinated	$(1 - v)a[d + (1 - d)k]\omega_U$	$(1 - v)(1 - k)(1 - ad\omega_U)$	$(1 - v)(1 - a\omega_U)k$

Contents of the table are probabilities of the various listed outcomes, accounting for vaccinated or unvaccinated status. Parameters: v , proportion of the population vaccinated; a , prevalence of SARS-CoV-2 infection; d , risk of COVID-19 among those with SARS-CoV-2 infection; k , risk of symptoms due to causes other than SARS-CoV-2 infection; θ_s , relative risk of SARS-CoV-2, given vaccination; θ_p , relative risk of COVID-19, among those with SARS-CoV-2 infection, given vaccination; β , relative risk of exposure or infection among vaccinated versus unvaccinated individuals attributable to factors other than vaccine-conferred immunity; ω_v , proportion of vaccinated individuals who remain uninfected; ω_U , proportion of unvaccinated individuals who remain uninfected (Table 1).

ACCEPTED

Figure 1

