

# Estimands and Estimation of COVID-19 Vaccine Effectiveness Under the Test-Negative Design

## Connections to Causal Inference

Mireille E. Schnitzer<sup>a,b,c</sup>

**Abstract:** The test-negative design is routinely used for the monitoring of seasonal flu vaccine effectiveness. More recently, it has become integral to the estimation of COVID-19 vaccine effectiveness, in particular for more severe disease outcomes. Because the design has many important advantages and is becoming a mainstay for monitoring postlicensure vaccine effectiveness, epidemiologists and biostatisticians may be interested in further understanding the effect measures being estimated in these studies and connections to causal effects. Logistic regression is typically applied to estimate the conditional risk ratio but relies on correct outcome model specification and may be biased in the presence of effect modification by a confounder. We give and justify an inverse probability of treatment weighting (IPTW) estimator for the marginal risk ratio, which is valid under effect modification. We use causal directed acyclic graphs, and counterfactual arguments under assumptions about no interference and partial interference to illustrate the connection between these statistical estimands and causal quantities. We conduct a simulation study to illustrate and confirm our derivations and to evaluate the performance of the estimators. We find that if the effectiveness of the vaccine varies across patient subgroups, the logistic regression can lead to misleading estimates, but the IPTW estimator can produce unbiased estimates. We also find that in the presence of partial interference both estimators can produce misleading estimates.

**Keywords:** Causal inference; Directed acyclic graphs; SARS-CoV-2; Test-negative design; Vaccine effectiveness

(*Epidemiology* 2022;33: 325–333)

Submitted September 28, 2021; accepted January 26, 2022

From the <sup>a</sup>Faculty of Pharmacy, Université de Montréal, Montréal, QC, Canada; <sup>b</sup>Department of Social and Preventive Medicine, School of Public Health, Université de Montréal, Montréal, QC, Canada; and <sup>c</sup>Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montréal, QC, Canada.

M.E.S. holds a Canada Research Chair from the Canadian Institutes of Health Research (CIHR) and receives research support from the Natural Sciences and Engineering Research Council of Canada (NSERC).

No human data was used in this methodological study. All code for simulated data generation is provided in the online repository <https://github.com/mirschn/TND>.

M.E.S. has recently received speaker fees from Biogen. She has no other real or perceived conflicts of interest.

Correspondence: Mireille E. Schnitzer, Faculty of Pharmacy, Université de Montréal, Montréal, QC, Canada. E-mail: [mireille.schnitzer@umontreal.ca](mailto:mireille.schnitzer@umontreal.ca).

Copyright © 2022 Wolters Kluwer Health, Inc. All rights reserved.

ISSN: 1044-3983/22/333-325

DOI: 10.1097/EDE.0000000000001470

The test-negative design is a type of observational study design routinely used to estimate seasonal influenza vaccine effectiveness.<sup>1,2</sup> It is currently being employed internationally to estimate coronavirus disease 2019 (COVID-19) vaccine effectiveness.<sup>3,4</sup> When prospectively implemented, this design recruits individuals who are seeking care or testing in response to COVID-like illness.<sup>5,6</sup> Participants are tested for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection, for example, by reverse transcription polymerase chain reaction test. The results of this test determine whether the participant is categorized as a COVID-positive “case” or a COVID-negative “control”. COVID-19 vaccination history and other patient information are then obtained, possibly from health records.<sup>5</sup> Another implementation of this design involves using electronic health data to retrospectively identify patients who sought care or testing due to COVID-like symptoms, their SARS-CoV-2 infection test results and vaccination status at the time of care-seeking, and demographic and clinical information about the patient.<sup>7,8</sup> In either version, vaccine effectiveness is typically estimated using a multivariable logistic regression of the test result (i.e., case status) conditional on vaccination status and measured confounders<sup>6–8</sup> but inverse probability of treatment weighting has also recently been used.<sup>5</sup>

Jackson and Nelson<sup>1</sup> provided the first formal framework for the test-negative design. Using contingency tables of a hypothetical population stratified by binary vaccination status, infection status, and binary propensity to seek care, they showed that case status odds-ratio estimands simplify to risk ratios for medically attended illness under certain assumptions.<sup>1</sup> Foppa et al.<sup>9</sup> justified the design and use of the case status odds ratio through mathematical models of infectious disease transmission. Several studies<sup>2,10,11</sup> used causal directed acyclic graphs (DAGs)<sup>12</sup> to explore different sources of bias that may manifest when estimating seasonal flu vaccine effectiveness with the test-negative design. Shi et al.<sup>11</sup> investigated bias of the standard test-negative design parameter with respect to the marginal risk ratio parameter under a multiplicative model with binary variables. Vandembroucke et al.<sup>13</sup> and Schnitzer et al.<sup>14</sup> investigated the identification of risk factors for COVID-19 disease and SARS-CoV-2 infection, respectively, under the test-negative design with

additional population controls. Lewnard et al.<sup>15</sup> reviewed the test-negative design in the COVID-19 context and proposed multiple strategies to limit bias due to confounding and misclassification of case status. Important insights arising from past literature include (1) Because only those seeking and accessing care can be enrolled in the study, the test-negative design can better control for confounding due to care-seeking behavior than case-control studies though residual bias is possible when care-seeking behavior is nonbinary.<sup>10,15</sup> However, a consequence is that vaccine effectiveness is only estimated in the subpopulation that has access to care.<sup>1,4</sup> (2) Because all subjects are tested for infection, the test-negative design may be less subject to measurement error than cohort studies<sup>1</sup>; however, there is also a danger of important bias when patients falsely testing negative for SARS-CoV-2 are considered to be “controls.”<sup>15</sup> (3) The test-negative design can only estimate vaccine effectiveness to prevent medically attended illness, such as illness leading to hospitalization. (4) Under this design, logistic regression is only valid when the vaccine has no effect on disease with similar presentations to the one being studied.<sup>1</sup>

Although many of the above-cited studies contributed to the theoretical justification of the test-negative design,<sup>1,9–11,15</sup> none formally derived the estimands under a statistical sampling framework such as the one used to justify the now-standard analytical methods used for case-control studies<sup>16</sup> or investigated connections to counterfactual causal parameters. Furthermore, no study has justified estimation with inverse probability of treatment weighting (IPTW) though at least one applied study has made use of this method.<sup>5</sup>

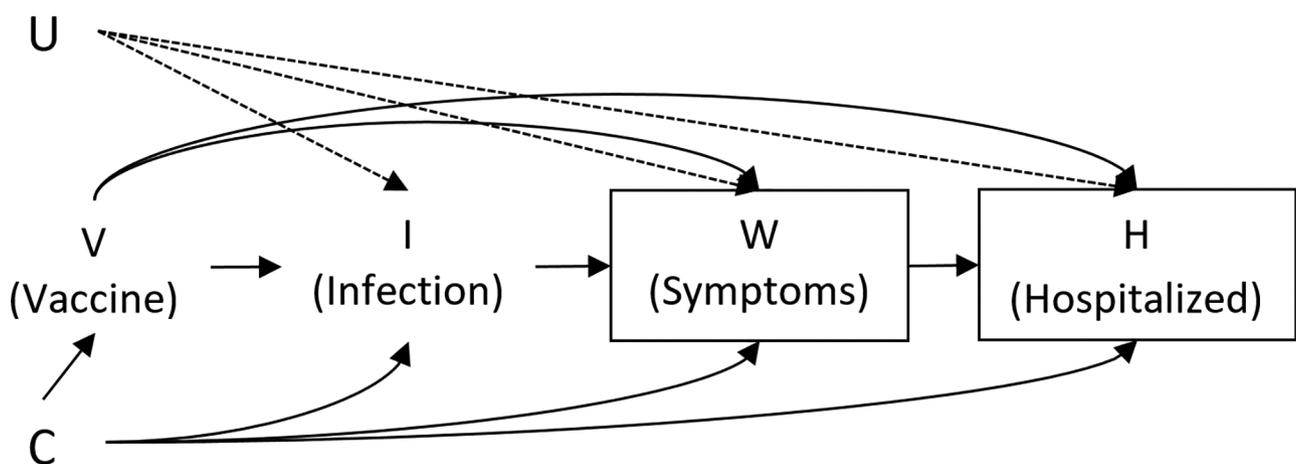
In this article, we postulate a nonparametric model, represented by a DAG, of the relevant variables at play when the

test-negative design is used to study vaccine effectiveness for medically attended COVID-19. Under this model, we derive the estimand of the test-negative design that is estimated with a correctly specified logistic regression. Under the assumption that the vaccine does not impact the probability of infection or disease due to another infection, the estimand is interpretable as an adjusted risk ratio for medically-attended COVID-19 with respect to vaccination status.<sup>1,11</sup> We also give the marginal risk ratio and show that it can be estimated using inverse probability of treatment weighting when the propensity score model (i.e., model for the probability of vaccination status) is fit using only the control data. In observational and experimental studies of vaccines for infectious disease, one person's disease occurrence may be impacted by the vaccination status of those in their entourage. This is called “interference” and complicates causal analysis.<sup>17</sup> We discuss potential connections of the statistical estimands to causal parameters under the assumptions of noninterference and partial interference, respectively. We conduct a simulation study to illustrate how estimates and estimands vary under effect modification and partial interference.

## STATISTICAL ESTIMANDS

We consider a test-negative design that recruits all patients admitted to hospital on the basis of specific symptoms of COVID-like illness which may be another infectious disease. The patients are then tested for SARS-CoV-2 infection. Administrative databases are then used to ascertain patients' history of COVID-19 vaccination.

The directed acyclic graph (DAG) in Figure represents a model of the progression of the variables considered in this study. First, individuals have some status of vaccination against COVID-19  $V$ , for example, “unvaccinated,”



**FIGURE.** Directed acyclic graph representing the hypothetical relationship between baseline confounders **C**, vaccination status  $V$ , infection  $I$  (none/SARS-CoV-2/other infection), severe symptoms  $W$ , hospitalization  $H$ , and unmeasured common causes of  $I$ ,  $W$ , and  $H$ . The variable  $U$  represents any (possibly unmeasured) common causes of infection, symptoms, and hospitalization that do not also affect vaccination. The boxes around  $W$  and  $H$  represent how the study sampling method conditions on these variables. Note that this conditioning also results in all participants having some kind of infection that can lead to severe COVID-19-like symptoms.

“fully vaccinated +  $x$  days”, etc. Subsequently, they may become infected with some virus ( $I$ ). Let  $I = 2$  denote infection by SARS-CoV-2,  $I = 1$  other infection, and  $I = 0$  the infection-free state. If an infection is present, the individual may develop severe symptoms  $W$ . These symptoms may lead to hospitalization  $H$  and thus inclusion in the study. Because this design is observational, there may be common causes  $C$  and  $U$  of any of the aforementioned variables, where  $U$  is unmeasured and assumed to not affect  $V$ . Confounders  $C$  may include age, comorbidities, employment sector, etc. The variable  $U$  could include latent COVID-19 susceptibility.

The test-negative design in this context presumably samples those with some infection ( $I \neq 0$ ), with severe symptoms ( $W = 1$ ) who are then hospitalized ( $H = 1$ ). Assuming a perfect test for SARS-CoV-2 infection, those who test positive,  $I = 2$ , have the outcome of interest and are considered cases. Those who test negative,  $I = 1$ , are considered controls. The index date of the study is the date of hospitalization and vaccination status is considered as of that date. This design is distinct from a standard case-control<sup>16</sup> because the participants are selected before knowledge of the nature of their infection.<sup>1</sup>

### The Conditional Risk Ratio Estimand

The observed data  $z = (c, v)$  are samples of  $Z = (C, V)$  from the probability function  $\Pr(Z = z | I \neq 0, W = 1, H = 1)$ . The infection type  $i = \{1, 2\}$  is sampled from a Bernoulli distribution with probability  $\Pr(I = i | Z = z, I \neq 0, W = 1, H = 1)$ . From this sampling, we can identify the odds ratio of data  $z$  versus some comparator  $z_0$  with respect to infection type by

$$\frac{\Pr(Z = z, I = 2 | I \neq 0, W = 1, H = 1)}{\Pr(Z = z_0, I = 2 | I \neq 0, W = 1, H = 1)} \bigg/ \frac{\Pr(Z = z, I = 1 | I \neq 0, W = 1, H = 1)}{\Pr(Z = z_0, I = 1 | I \neq 0, W = 1, H = 1)} \quad (1)$$

$$= \frac{\Pr(Z = z | I = 2, W = 1, H = 1)}{\Pr(Z = z_0 | I = 2, W = 1, H = 1)} \bigg/ \frac{\Pr(Z = z | I = 1, W = 1, H = 1)}{\Pr(Z = z_0 | I = 1, W = 1, H = 1)}$$

For example, we may contrast those fully vaccinated  $z = (c, v = \text{"fully vaccinated + 14 days"})$  vs. those unvaccinated  $z_0 = (c, v = \text{"unvaccinated"})$ . By the identity

$$\Pr(I = i, W = 1, H = 1 | Z = z) = \frac{\Pr(Z = z, I = i | I \neq 0, W = 1, H = 1) \Pr(I \neq 0, W = 1, H = 1)}{\Pr(Z = z)}$$

for  $i = 1, 2$  we can rewrite Equation (1) as

$$\frac{\Pr(I = 2, W = 1, H = 1 | Z = z)}{\Pr(I = 1, W = 1, H = 1 | Z = z)} \bigg/ \frac{\Pr(I = 2, W = 1, H = 1 | Z = z_0)}{\Pr(I = 1, W = 1, H = 1 | Z = z_0)} \quad (2)$$

which represents the prospective adjusted odds ratio of being hospitalized for COVID-19 vs. being hospitalized for another infection between fully vaccinated and unvaccinated individuals. Under the assumption that the COVID-19 vaccine has no impact on any other type of infection, the ensuing disease severity and potential for hospitalization,

$\Pr(I = 1, W = 1, H = 1 | Z = z) = \Pr(I = 1, W = 1, H = 1 | Z = z_0)$  and the odds ratio in Equation (2) simplifies to

$$\psi_{\text{cRR}} = \frac{\Pr(I = 2, W = 1, H = 1 | Z = z)}{\Pr(I = 2, W = 1, H = 1 | Z = z_0)}, \quad (3)$$

which is the prospective adjusted risk ratio of hospitalization for COVID-19 between fully vaccinated and unvaccinated individuals. The “vaccine effectiveness” estimand is typically given as  $1 - \psi_{\text{cRR}}$ .

Now applying the identity

$$\Pr(I = i, W = 1, H = 1 | Z = z) = \Pr(I = i | Z = z, I \neq 0, W = 1, H = 1) \Pr(I \neq 0, W = 1, H = 1 | Z = z)$$

we can rewrite Equation (2) as

$$\frac{\Pr(I = 2 | Z = z, I \neq 0, W = 1, H = 1)}{\Pr(I = 1 | Z = z, I \neq 0, W = 1, H = 1)} \bigg/ \frac{\Pr(I = 2 | Z = z_0, I \neq 0, W = 1, H = 1)}{\Pr(I = 1 | Z = z_0, I \neq 0, W = 1, H = 1)} \quad (4)$$

A multivariable logistic regression of  $I = 2$  vs.  $I = 1$ , conditional on  $C$  and a factorization of  $V$  using the data sampled under the test-negative design will yield estimates of the odds ratios in Equation (4) under the assumption that the logistic regression model for  $\Pr(I = i | Z = z, I \neq 0, W = 1, H = 1)$  is correctly specified.<sup>16</sup> Because of the equivalence of Equation (4) with Equation (3), we can say that the exponential of the coefficient related to vaccination status in the logistic regression is an estimate of the adjusted risk ratio for hospitalization with COVID-19.

Thus, under the assumed DAG and logistic regression model, the conditional risk ratio in Equation (3) is the interpretable estimand. Importantly, this risk ratio is an association between vaccination and a combined outcome that involves three steps: becoming infected with SARS-CoV-2, having severe symptoms, and accessing (being admitted to) a hospital due to these symptoms.<sup>1</sup>

### The Marginal Risk Ratio

The risk ratio is collapsible, so its value does not depend on the variables in the conditioning set beyond adjustment for confounding. But estimation using logistic regression relies on correct model specification, which may be implausible in practice. In particular, if vaccine effectiveness differs by subgroup, then estimating an overall effect with logistic regression will inevitably result in misspecification because the model cannot include interactions with vaccination. One may choose to instead directly estimate the marginal risk ratio.

$$\psi_{\text{mRR}} = \frac{\mathbb{E}\{\Pr(I = 2, W = 1, H = 1 | Z = z)\}}{\mathbb{E}\{\Pr(I = 2, W = 1, H = 1 | Z = z_0)\}} \quad (5)$$

Under effect modification where a third variable can change the effect of vaccination on infection, disease, or hospitalization, this estimand is not equal to the conditional risk ratio.

Due to the biased sampling design, one cannot directly estimate the conditional probability of experiencing an outcome. However, under the previous assumption that the vaccine has no impact on other infections, we have that

$$\begin{aligned} & \Pr(V = v \mid I = 1, W = 1, H = 1, C = c) \\ &= \frac{\Pr(I = 1, W = 1, H = 1, Z = z \mid V = v, C = c) \Pr(V = v \mid C = c)}{\Pr(I = 1, W = 1, H = 1, Z = z \mid C = c)} \\ &= \Pr(V = v \mid C) \end{aligned}$$

This means that we can estimate the propensity score by (1) modeling the conditional probability of vaccine status using only the controls, then (2) using this model to estimate probabilities for the whole sample.

It is then possible to use IPTW to estimate the marginal risk ratio directly. This involves taking a ratio of weighted means between subjects with different vaccination statuses. Consider a sample of  $n$  hospitalized patients with observed data  $(c_k, v_k, i_k; k = 1, \dots, n)$  corresponding to measured confounders, vaccination status, and SARS-CoV-2 infection status, respectively. The latter IPTW estimator contrasting vaccination status  $v$  vs.  $v_0$  can be written as follows:

$$\frac{1}{n} \sum_{k=1}^n \frac{\mathbb{I}(i_k = 2 \text{ and } v_k = v)}{\widehat{\Pr}(V = v \mid C = c_k)} / \frac{1}{n} \sum_{k=1}^n \frac{\mathbb{I}(i_k = 2 \text{ and } v_k = v_0)}{\widehat{\Pr}(V = v_0 \mid C = c_k)},$$

where  $\mathbb{I}(\cdot)$  represents the indicator function and  $\widehat{\Pr}(V = v_0 \mid C = c_k)$  represents the estimate of the propensity score for the  $k$ th patient. One could alternatively run a weighted logistic regression. Estimation with IPTW in the biased sample is justified theoretically in the Appendix and empirically in the simulation study.

### CAUSAL INTERPRETATION

Whether statistical estimands can be connected to causal estimands depends on additional assumptions. Causal estimands are often defined based on potential outcomes. A potential outcome under a given treatment is the outcome a participant would have had if they had received that treatment, for example, the outcome they would have under a given vaccination status. Causal inference operating under Rubin’s “stable unit treatment value assumption”<sup>18</sup> requires that the treatment (vaccination) of one individual does not impact the potential outcome of another. This assumption, also referred to as the absence of interference, is likely violated in studies of COVID-19 vaccination effectiveness due to widespread vaccination programs and reduced infectiousness among fully vaccinated people.<sup>17,19</sup> We explore the relationships between the statistical estimands defined above and causal estimands

under the hypothetical assumptions of no interference and partial interference.

### Causal Interpretation Assuming No Interference

In the absence of interference, one patient’s infection, illness, and hospitalization under their vaccination status could not depend on another patient’s vaccination status. We would then assume consistency where assignment to a vaccination status  $V = v$  yields the same outcome as when the observed vaccination status is  $V = v$ . If we define the potential outcome under vaccination status  $v$  as  $\{I(v), W(v), H(v)\}$ , this means that  $\{I, W, H\} = \{I(v), W(v), H(v)\}$  when  $V = v$ . Under the DAG in the figure, we must assume that we have measured all variables  $C$ . Specifically, we must satisfy the conditional ignorability assumption  $\{I(v), W(v), H(v)\} \perp\!\!\!\perp V \mid C$ . This also means that any common cause of vaccination status with any of the three outcome variables must be measured.

In addition, the probability of an individual having any vaccination status must be non-zero for any possible value of  $C$ . This last assumption would be violated if, for instance, the study covers a time-period where some of the recruited patients could not have possibly been fully vaccinated due to age-restrictions during the roll-out period.

Supposing that the above assumptions were true, the risk ratio in equation (3) contrasting some vaccination statuses  $v = 1$  versus  $v = 0$  could be rewritten as

$$\psi_{\text{CausalRR}} = \frac{\Pr(I(1) = 2, W(1) = 1, H(1) = 1 \mid C = c)}{\Pr(I(0) = 2, W(0) = 1, H(0) = 1 \mid C = c)} \quad (6)$$

which is the conditional causal risk ratio of being hospitalized with COVID-19. The marginal risk ratio can be rewritten as

$$\psi_{\text{mCausalRR}} = \frac{\Pr(I(1) = 2, W(1) = 1, H(1) = 1)}{\Pr(I(0) = 2, W(0) = 1, H(0) = 1)} \quad (7)$$

However, given that interference is likely in vaccine studies of infectious diseases, these causal parameters may not be well-defined.

### Causal Interpretation Under Interference

Under general interference, one person’s potential outcome, indexed by  $j$ , depends on the full vector of other individuals’ treatment statuses,  $\mathbf{v}_{-j} = (v_1, v_2, \dots, v_{j-1}, v_{j+1}, \dots, v_n)$ . This individual’s potential outcome can be denoted  $Y_j(v_j, \mathbf{v}_{-j})$ . If interference is limited to known blocks or networks, causal effects can be defined and potentially estimated. For example, in a multisite study, suppose that participants are geographically connected by site so that interference only exists between individuals accessing care at a common site. We say that these individuals are in the same “block.”<sup>20</sup> Let  $\mathbf{v}^m$  denote a vector of vaccination statuses for the  $m$ th block and let  $\mathbf{v}_{-j}^m$  denote the same vector with the  $j$ th element removed. Suppose that

we can define the potential outcome of the  $j$ th individual in block  $m$  as depending on a one-dimensional summary  $f(\mathbf{v}_{-j}^m)$  of the block's overall vaccine uptake (called "vaccination coverage").<sup>21</sup> The potential outcome for individual  $j$  in block  $m$  could then be rewritten as  $Y_j(v_j, \mathbf{v}_{-j}) = Y_j(v_j, f(\mathbf{v}_{-j}^m))$ . Given two block vaccination coverage levels  $f$  and  $f'$  and two individual vaccination statuses  $v$  and  $v'$ , estimands of interest in this framework include the direct effect  $\mathbb{E}\{Y(v, f)\} / \mathbb{E}\{Y(v', f)\}$ , which is the effect of changing an individual's vaccination status but fixing the vaccination status of others in the block; the indirect effect  $\mathbb{E}\{Y(v, f)\} / \mathbb{E}\{Y(v, f')\}$ , which is the effect of changing the block's vaccination coverage when the individual's vaccination status is held fixed; and the total effect  $\mathbb{E}\{Y(v, f)\} / \mathbb{E}\{Y(v', f')\}$  of changing both the individual vaccination status and block's vaccination coverage.<sup>17</sup>

In standard contexts, these causal estimands can be estimated under challenging assumptions about measured confounding.<sup>20</sup> First, all confounders of individual vaccination and outcome must be adjusted for. But because we also need to unconfound the relationship between the individual outcome and the block's vaccination coverage, it may also be necessary to adjust for summaries of the block's covariates.<sup>22</sup> Examples of such covariates are summaries of the political orientations of the block's members and leadership which can impact vaccination uptake and are related to health risk-taking behavior. Let  $\mathbf{c}_{-j}^m$  denote the vector of all covariates in block  $m$  except individual  $j$ . The block-level summary can be denoted  $\mathbf{g}_m = \mathbf{g}(\mathbf{c}_{-j}^m)$ . If these measured variables allow for the adjustment for confounding, then we have that

$$E(Y | C = c, G = g(c), V = v, F = f) = E(Y(v, f) | C = c, G = g(c)).$$

In the test negative design, the outcome for individual  $j$  in block  $m$  is  $Y_j = \{I_j, W_j, H_j\}$  which we assume is equal to the potential outcomes  $\{I_j(v, f_m), W_j(v, f_m), H_j(v, f_m)\}$  when we observe  $V_j = v$  and  $f(\mathbf{V}_{-j}^m) = f_m$ . And so we can write the conditional risk ratio estimated by the standard logistic regression as

$$\begin{aligned} \psi_{cRR} &= \frac{\Pr(I = 2, W = 1, H = 1 | C = c, G = g(C), V = v)}{\Pr(I = 2, W = 1, H = 1 | C = c, G = g(C), V = v')} \\ &= \frac{\sum_{m=1}^M \Pr(S_m) \mathbb{E}_{P_m}(Y | C = c, G = g_m, V = v, F = f_m)}{\sum_{m=1}^M \Pr(S_m) \mathbb{E}_{P_m}(Y | C = c, G = g_m, V = v', F = f_m)}, \end{aligned}$$

where  $\Pr(S_m)$  is the proportion of the source population in block  $m$  and the second-line expectations are taken under  $P_m$ , the block-specific density function. With no unmeasured confounders,  $\psi_{cRR}$  is equal to

$$\frac{\sum_{m=1}^M \Pr(S_m) \Pr_{P_m}(I(v, f_m) = 2, W(v, f_m) = 1, H(v, f_m) = 1 | C = c, G = g_m)}{\sum_{m=1}^M \Pr(S_m) \Pr_{P_m}(I(v', f_m) = 2, W(v', f_m) = 1, H(v', f_m) = 1 | C = c, G = g_m)}$$

So if block-level confounder summaries are adjusted for in the analysis and if those and the individual-level confounders are sufficient to identify block-specific causal effects, then the overall conditional risk ratio estimand can roughly be interpreted as a contrast between the weighted averages of the block-specific probabilities of medically-attended disease. However, because  $f_m$  is the observed vaccination coverage in block  $m$  and because this coverage can vary by block, this is not a causal contrast in the sense that it does not represent a marginal or conditional effect of intervening on vaccination. In the simulation study, we explore the potential deviation between the estimates from the test-negative design and the causally-interpretable direct effect risk ratio

$$\psi_{cRR, f^*} = \frac{\Pr(I(v, f^*) = 2, W(v, f^*) = 1, H(v, f^*) = 1 | C = c, G = g(C))}{\Pr(I(v', f^*) = 2, W(v', f^*) = 1, H(v', f^*) = 1 | C = c, G = g(C))}$$

under vaccination coverage  $f^*$ .

### SIMULATION STUDIES

The objective of the simulation study is to compare the values of statistical and causal estimands and to evaluate estimation by logistic regression and the proposed IPTW in the test-negative design. We generated three scenarios, summarized in Table 1. The first two followed the DAG in the Figure with single continuous confounder  $C$  and all other variables binary. Two binary unmeasured variables impacted the risks of SARS-CoV-2 infection, severe symptoms due to COVID-19, and hospitalization if the severe symptoms were present. The second unmeasured variable also impacted the risk of other infection and severe symptoms due to other disease. Severe disease symptoms were generated separately for infection with SARS-CoV-2 and other infection. Hospitalization was only possible with severe disease. Vaccination offered protection from infection, severe disease, and hospitalization. For the first two scenarios, no interference was present.

- Scenario 1 (Basic Setting): No effect modification of vaccination  $V$  by covariate  $C$  in any model, no effect of the vaccine on non-SARS-CoV-2 infection or disease. The effects of the vaccine  $V$  and symptoms  $W$  on hospitalization did not depend on infection type.
- Scenario 2 (Effect modification): Same as Scenario 1 except with an added interaction term between  $C$  and  $V$  in the model for infection with SARS-CoV-2. This represents a different individual effect depending on the patient's value of  $C$ . For example, if  $C$  represents age, then this represents a scenario where older people do not benefit as much from the vaccine as younger people.

The third scenario introduced partial interference.

- Scenario 3 (Partial interference): The subjects in the simulation belonged to 10 blocks of fixed sizes. We generated three subject-level covariates: one continuous and measured, and two binary and unmeasured.

**TABLE 1.** Data Generation Structure in the Simulation Study

Variable Name and Type	Description	Generated Conditional On
<i>Scenarios 1 and 2: generated from single population.</i>		
<i>Scenario 3: generated in 10 blocks of fixed size.</i>		
$C$ , conts	Baseline confounder	None
(Scenario 3 only:)	Study-level baseline confounder (common value within	None
$X$ , conts	block), e.g., local incidence	
$U_1$ and $U_2$ , bin	Unmeasured covariates (causes of outcome)	None
$V$ , bin	Vaccination	Scenario 3: $U_2$ depends on $X$
$I_1 = 1$ , bin	Infection with other virus	$C$
$I_2 = 1$ , bin	Infection with SARS-CoV-2	Scenario 3: $+X$ and block
		$C, U_1$
		$C, V, U_1, U_2, (1 - V) * U_2$
		Scenario 2: $+V * C$
		Scenario 3: $+X$
$W[I = 1]$ , bin	Severe disease due to other virus	$C, V, U_1$
$W[I = 2]$ , bin	Severe COVID-19	$C, V, U_1,$
		Scenarios 1 and 2: $+(1 - V) * U_2$
		Scenario 3: $+U_2$ and interaction between $V$ and % vaccinated in block
$H[W = 1]$ , bin	Hospitalization with severe symptoms	$C, V, U_1$
		Scenario 3: $+%$ vaccinated in block

*Study sample, scenarios 1 and 2: randomly sample  $n$  patients with  $H = 1$*

*Study sample, scenario 3: randomly sample  $n = 10$  patients with  $H = 1$  per block*

Each given variable is univariate and generated randomly conditional on the variables given in the rightmost column. bin indicates binary (generated as Bernoulli), conts: continuous (generated as Gaussian).

We also generated a continuous block-level covariate,  $X$ , representing a measure of local incidence of infection with SARS-CoV-2. The probability of vaccination depended on the block, and the local incidence  $X$  where greater incidence encouraged vaccination. Infection with SARS-CoV-2 was affected by vaccination status, the block- and individual-level covariates, and the overall vaccine uptake (proportion vaccinated) in the rest of the block. The proportion vaccinated was included as an effect modifier of vaccination with the hypothesis that there is less exposure to the virus when more surrounding people are vaccinated, and that vaccination in addition to lower viral exposure is more protective than the sum of each individual element.

In all scenarios, logistic regression and IPTW adjusted for the baseline confounder  $C$ . In the third scenario, the models also adjusted for the block-level covariate  $X$ .

The results are presented in Table 2. In all scenarios, the vaccine is more effective at reducing hospitalization than infection. When there was no interference, both the conditional ( $\psi_{cRR}$ ) and marginal ( $\psi_{mRR}$ ) risk ratio estimands have a causal interpretation, corresponding to the counterfactual parameters  $\psi_{cCausalRR}$  and  $\psi_{mCausalRR}$ , respectively.

In the first scenario, there was no effect modification so the conditional and marginal risk ratios were equal. The effect of vaccination on infection with SARS-CoV-2 (0.84) was lower than for hospitalization with COVID-19 (0.96). In terms of

estimation, the logistic regression performed well although standard confidence intervals undercovered the true conditional causal effect. IPTW performed better with higher coverage of the marginal causal effect when the propensity score was estimated using the controls. The IPTW with propensity score estimated using all of the data was highly biased in all scenarios.

In the second scenario, effect modification resulted in subpopulation vaccine effectiveness that differed by value of  $C$ : vaccine effectiveness was 0.98 in the first three quartiles of  $C$ , 0.70 in the fourth quartile, and only 0.56 in the 95th percentile. This is akin to vaccine effectiveness dropping off substantially only for the elderly. The overall marginal and conditional vaccine effectiveness were slightly different (0.75 and 0.77, respectively). Logistic regression incorrectly averaged over subgroup effects, resulting in a large overestimate of vaccine effectiveness. When we stratified the logistic regression on subjects with  $C$  values in the fourth quartile, we obtained a mean estimate of 0.84, which was also biased for the vaccine effectiveness in that quartile (0.70). IPTW with the control-estimated propensity score performed very well for the estimation of the marginal causal effect. Logistic regression had a lower variance than IPTW because the former ignores the variability in vaccine effectiveness across members of the population.

In the third scenario, we calculated the true conditional and marginal risk ratios. Because interference was present, these no longer have a causal interpretation. We also

**TABLE 2.** Simulation Study Results

	Truth	Mean Est	MC SE	% Cov <sub>cRR</sub>	% Cov <sub>mRR</sub>
<i>Scenario 1: Basic setting</i>					
<b>Estimands</b>					
$1 - \psi_{cRR}$	0.96				
$1 - \psi_{mRR}$	0.96				
$1 - mRR$ for SARS-CoV-2	0.84				
<b>Estimators</b>					
Logistic regression		0.97	0.01	89	-
IPTW					
$\pi_{controls}$		0.96	0.02	-	92
$\pi_{all}$		0.44	0.06	-	0
<i>Scenario 2: Effect modification by C for SARS-CoV-2 infection</i>					
<b>Estimands</b>					
$1 - \psi_{cRR}$	0.77				
$1 - \psi_{mRR}$	0.75				
$1 - mRR$ for SARS-CoV-2	0.44				
<b>Estimators</b>					
Logistic regression		0.91	0.06	39	-
IPTW					
$\pi_{controls}$		0.74	0.15	-	93
$\pi_{all}$		0.18	0.05	-	0
<i>Scenario 3: Partial interference by block vaccination prevalence, f</i>					
<b>Estimands</b>					
$1 - \psi_{cRR}$	0.86				
$1 - \psi_{mRR}$	0.85				
$1 - mRR$ for SARS-CoV-2	0.48				
$1 - \psi_{cRR,75}$	0.86				
$1 - \psi_{cRR,50}$	0.80				
$1 - \psi_{cRR,25}$	0.76				
<b>Estimators</b>					
Logistic regression		0.88	0.04	89	-
IPTW					
$\pi_{controls}$		0.85	0.06	-	92
$\pi_{all}$		0.14	0.04	-	0

Aggregate results of the application of each method to 1,000 simulated datasets of  $n$  hospitalized patients where  $n = 500$  for Scenarios 1 and 2 and  $n = 1000$  for Scenario 3. The results are given with respect to one minus the risk ratios, often referred to as “vaccine effectiveness.”  $\psi_{cRR}$ : the conditional risk ratio for hospitalization with COVID-19 in Equation (3);  $\psi_{mRR}$ : the marginal risk ratio for hospitalization with COVID-19 in Equation (5). % Cov indicates % of 95% confidence intervals that contain the true vaccine effectiveness (optimal is 95%); Mean est, mean estimate; MC SE, Monte-Carlo standard error of the estimate;  $mRR$  marginal risk ratio.

computed the direct effect of vaccination  $1 - \psi_{cRR,f}$ , defined as the risk ratio of hospitalization with COVID-19 contrasting vaccination status and fixing the proportion vaccinated in a block ( $f$ ). We give this estimand for three values of  $f = 0.75, 0.5, 0.25$ . The impact of vaccination was the highest (0.86) when 75% of a block was vaccinated. The impact was lower when half or a quarter of the block was vaccinated, with direct effects of 0.80, and 0.76, respectively. In this scenario, logistic regression slightly overestimated the conditional association between the vaccine and hospitalization with COVID-19. IPTW accurately estimated the marginal association. Both estimators gave misleading estimates if they were to be interpreted in a population with lower vaccination coverage.

Under partial interference by block, one may stratify estimation by block to directly estimate the causal parameters  $\psi_{cRR,f_m}$  for the block-specific vaccination prevalence value  $f_m$ . Table 3 compares the results of pooled and block-stratified estimation for a single simulated dataset, representing a census of the hospitalized patients (note that in the previous analysis, we took a random subset of patients from each block to represent participants in the study). The analyses of the pooled data had a large sample size of 13,731, and both estimates produced little error relative to the true risk ratios. However, the stratified estimators had smaller sample sizes (particularly for controls) leading to much greater error in the estimates, in particular for the studies with lower vaccination prevalence.

**TABLE 3.** Single Simulated Dataset From Scenario 3: Block-Stratified and Pooled Estimation of Vaccine Effectiveness With Logistic Regression and IPTW

Block	Population Size	Sample Size (n. Controls)	Population Vaccination Prevalence, $f^*$	True Values $1 - \psi_{cRR}, f^*$	Logistic Regression, Est (95% CI)	IPTW, Controls, Est (95% CI)
2	250,000	1,832 (31)	0.24	0.76	0.44 (−0.70, 0.77)	0.0 (−2.2, 0.65)
3	250,000	895 (26)	0.32	0.77	0.73 (0.35, 0.88)	0.64 (0.17, 0.84)
1	250,000	523 (48)	0.37	0.77	0.69 (0.38, 0.84)	0.66 (0.34, 0.82)
5	500,000	1,665 (73)	0.58	0.82	0.86 (0.77, 0.92)	0.86 (0.75, 0.92)
4	250,000	346 (38)	0.69	0.84	0.89 (0.77, 0.96)	0.92 (0.81, 0.96)
8	500,000	1,578 (78)	0.75	0.86	0.84 (0.73, 0.91)	0.79 (0.64, 0.88)
10	1,000,000	2,816 (134)	0.77	0.87	0.93 (0.88, 0.96)	0.90 (0.84, 0.94)
9	1,000,000	2,379 (147)	0.78	0.87	0.89 (0.83, 0.93)	0.85 (0.76, 0.90)
6	500,000	1,121 (78)	0.79	0.87	0.88 (0.78, 0.93)	0.84 (0.71, 0.91)
7	500,000	576 (77)	0.83	0.88	0.91 (0.82, 0.96)	0.87 (0.74, 0.94)
Analysis of pooled data, n = 13,731:			Adjusting for $X$		0.87 (0.84, 0.89)	0.85 (0.81, 0.87)
			Adjusting for block		0.86 (0.84, 0.87)	0.83 (0.80, 0.86)

This analysis was conducted on a single simulated dataset representing a census of hospitalized patients, allowing for a larger sample size in each block. IPTW was implemented with weighted logistic regression where only controls were used to fit the propensity score model. The stratified analyses used 10% weight truncation, but this had negligible impact on the estimates of all blocks except for block 2 where the estimation was unstable due to only having five vaccinated controls.

## DISCUSSION

The test-negative design is being increasingly used to study postlicensure vaccine effectiveness for COVID-19.<sup>4</sup> In this article, we placed the design in a causal inference context by presenting a nonparametric model related to vaccination for SARS-CoV-2, infection, disease symptoms, and reception of care, such as hospitalization. Under this model, we derived the identifiable conditional risk ratio estimand under the statistical sampling framework that differentiated between inclusion criteria (symptoms and accessing care) and measured data (covariates, vaccination status, and infection status). We demonstrated that an IPTW estimator for the risk ratio can be implemented when the propensity score model is fit using only control data. This approach has also been used for case-control studies<sup>23,24</sup> where the validity of the propensity score estimation relies instead on a rare disease assumption. In the test-negative design, neither estimator requires a rare disease assumption. However, both estimators require that vaccination for COVID-19 has no impact on other diseases with similar symptoms.

The major benefit of IPTW is that, unlike logistic regression, it estimates marginal effects even when effect modification is present. While one may present stratified estimates of vaccine effectiveness for different age subgroups, for instance, effect modification is likely to be high-dimensional, also depending on comorbidities and immunosuppressant drug usage. Thus there is likely effect modification occurring even within age groups, leading to potential bias in the logistic regression estimator. IPTW avoids the issue of needing to stratify on a subset where there is no residual effect modification. This also avoids having only small sample sizes to estimate the effects of interest.

In vaccine studies, including randomized controlled trials, interference typically complicates the definition and

estimation of causal estimands.<sup>17</sup> In our discussion on interference, we made the strong assumption that one person's outcome (infection, severe symptoms, and hospitalization) can only be affected by their block's vaccination coverage, and not the vaccination statuses of the individuals inside or outside of the block. A block may be approximately defined by the geographic region of the medical facility where the patient was admitted.<sup>5</sup> If all confounders of individual vaccination and block vaccination coverage with the individual's outcome are measured, then the conditional risk ratio is a ratio of weighted sums of block-specific causal risks. This means that we are contrasting averages of block-vaccination-rate-specific risks under (counterfactual) assignment of vaccination status to the individual. In the simulation study, we presented a scenario where the conditional and marginal risk ratios represented the impact of vaccination only under high vaccination coverage. If we were to interpret either risk ratio as the impact of vaccination, we would be overestimating the vaccination effectiveness except when vaccination coverage is high. Note that this is one specific scenario but it illustrates that average associations over populations with heterogeneous vaccination coverage can be misleading and not represent a causal effect. We also showed that when presenting results by block,<sup>5</sup> we target a causally interpretable block-stratified parameter, but risk greater estimation error due to smaller sample sizes.

Similarly, we could repeat this exercise by considering interference and effect-modification by local infection rates. The noted challenges point to the importance of stratifying the analysis by block when sample sizes are sufficiently large and developing valid pooled estimators under various types of interference for this design.

Although the test-negative design is not new, emerging evidence of COVID-19 vaccination effectiveness has put it in

the international spotlight. Careful study of this design and the interpretation of estimates obtained under this design are ongoing research priorities.

REFERENCES

1. Jackson ML, Nelson JC. The test-negative design for estimating influenza vaccine effectiveness. *Vaccine*. 2013;31:2165–2168.
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;45:2060–2074.
3. Patel MM, Jackson ML, Ferdinands J. Postlicensure evaluation of COVID-19 vaccines. *JAMA*. 2020;324:1939–1940.
4. Dean NE, Hogan JW, Schnitzer ME. Covid-19 vaccine effectiveness and the test-negative design. *N Engl J Med*. 2021;385:1431–1433.
5. Thompson MG, Stenehjem E, Grannis S, et al. Effectiveness of COVID-19 vaccines in ambulatory and inpatient care settings. *N Engl J Med*. 2021;385:1355–1371.
6. Hyams C, Marlow R, Maseko Z, et al. Effectiveness of BNT162b2 and ChAdOx1 nCoV-19 COVID-19 vaccination at preventing hospitalisations in people aged at least 80 years: a test-negative, case-control study. *Lancet Infect Dis*. 2021;21:1539–1548.
7. Chung H, He S, Nasreen S, et al. Effectiveness of BNT162b2 and mRNA-1273 covid-19 vaccines against symptomatic SARS-CoV-2 infection and severe covid-19 outcomes in Ontario, Canada: test negative design study. *BMJ*. 2021;374:n1943.
8. Lopez Bernal J, Andrews N, Gower C, et al. Effectiveness of the Pfizer-BioNTech and Oxford-AstraZeneca vaccines on covid-19 related symptoms, hospital admissions, and mortality in older adults in England: test negative case-control study. *BMJ*. 2021;373:n1088.
9. Foppa IM, Haber M, Ferdinands JM, Shay DK. The case test-negative design for studies of the effectiveness of influenza vaccine. *Vaccine*. 2013;31:3104–3109.
10. Sullivan SG, Tchetgen Tchetgen EJ, Cowling BJ. Theoretical basis of the test-negative study design for assessment of influenza vaccine effectiveness. *Am J Epidemiol*. 2016;184:345–353.
11. Shi M, An Q, Ainslie K, Haber M, Orenstein WA. A comparison of the test-negative and the traditional case-control study designs for estimation of influenza vaccine effectiveness under nonrandom vaccination. *BMC Infect Dis*. 2017;17:757.
12. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology*. 1999;10:37–48.
13. Vandembroucke JP, Brückley EB, Vandembroucke-Grauls CMJE, Pearce N. The test-negative design with additional population controls: a practical approach to rapidly obtain information on the causes of the SARS-CoV-2 epidemic. *Epidemiology*. 2020;31:836–843.
14. Schnitzer ME, Harel D, Ho V, Koushik A, Merckx J. Identifiability and estimation under the test-negative design with population controls with the goal of identifying risk and preventive factors for SARS-CoV-2 infection. *Epidemiology*. 2021;32:690–697.
15. Lewnard JA, Patel MM, Jewell NP, et al. Theoretical framework for retrospective studies of the effectiveness of SARS-CoV-2 vaccines. *Epidemiology*. 2021;32:508–517.
16. Prentice RL, Pyke R. Logistic disease incidence models and case-control studies. *Biometrika*. 1979;66:403–411.
17. Hudgens MG, Halloran ME. Toward causal inference with interference. *J Am Stat Assoc*. 2008;103:832–842.
18. Rubin DB. Statistics and causal inference: which ifs have causal answers. *J Am Stat Soc*. 1986;81:961–962.
19. Centers for Disease Control and Prevention. Science Brief: COVID-19 Vaccines and Vaccination. Available at: <https://www.cdc.gov/coronavirus/2019-ncov/science/science-briefs/fully-vaccinated-people.html>. 2021. Accessed 9 September 2021.
20. Ogburn EL, VanderWeele TJ. Causal diagrams for interference. *Stat Sci*. 2014;29:559–578.

21. Hong G, Raudenbush SW. Evaluating kindergarten retention policy: a case study of causal inference for multi-level observational data. *J Am Stat Assoc*. 2006;101:901–910.
22. VanderWeele TJ, Tchetgen Tchetgen EJ, Halloran ME. Interference and sensitivity analysis. *Stat Sci*. 2014;29:687–706.
23. Robins JM. [Choice as an alternative to control in observational studies]: comment. *Stat Sci*. 1999;14:281–293.
24. Månsson R, Joffe MM, Sun W, Hennessy S. On the estimation and use of propensity scores in case-control and case-cohort studies. *Am J Epidemiol*. 2007;166:332–339.

APPENDIX: PROOF OF CONSISTENCY FOR THE IPTW ESTIMATOR

**Claim:** Let  $\mathbb{E}_{TND}$  represent an expectation under test-negative design biased sampling and  $\mathbb{E}$  be an expectation under unbiased sampling. Let  $Y = \mathbb{I}(I = 2, W = 1, H = 1)$ , the outcome of interest. Let  $q_0 = \Pr(I \neq 0, W = 1, H = 1)$  be the marginal probability of having the inclusion criteria for the study. Then, we have that

$$\mathbb{E}_{TND} \left\{ \frac{Y \mathbb{I}(V = v) q_0}{\Pr(V = v | C = c)} \right\} = \mathbb{E} \{ \Pr(I = 2, W = 1, H = 1 | V = v, C = c) \}.$$

*Proof.* Let  $pr_C(c)$  and  $p_C(c | I \neq 0, W = 1, H = 1)$  be the probability density functions of the covariates  $C$  under simple random sampling and test-negative design sampling, respectively. Also let  $S = \mathbb{I}(I \neq 0, W = 1, H = 1)$  indicate the presence of inclusion criteria for the test-negative design.

$$\begin{aligned} \mathbb{E}_{TND} \left\{ \frac{Y \mathbb{I}(V = v) q_0}{\Pr(V = v | C = c)} \right\} &= \int_c \frac{Y \mathbb{I}(V = v) q_0}{\Pr(V = v | C = c)} p_C(c | S = 1) dc \\ &= \int_c \frac{\Pr(Y = 1, V = v | S = 1, C = c)}{\Pr(V = v | C = c)} \Pr(S = 1 | C = c) p_C(c) dc \\ &= \int_c \frac{\Pr(Y = 1, V = v, S = 1 | C = c)}{\Pr(V = v | C = c)} p_C(c) dc \\ &= \int_c \Pr(Y = 1, S = 1 | V = v, C = c) p_C(c) dc \\ &= \int_c \Pr(Y = 1 | V = v, C = c) p_C(c) dc \\ &= \mathbb{E} \{ \Pr(I = 2, W = 1, H = 1 | V = v, C = c) \}. \end{aligned}$$

Thus, if we weight observations by the inverse of  $\Pr(V = v | C = c)$  (which, as we recall, must be modeled using only the control data), we can recover the marginal mean of the outcome under a vaccination status  $v$  up to the constant  $q_0$ . Therefore, one can only estimate the marginal probability of the outcome (i.e., the numerator of Equation 5) with knowledge of  $q_0$ . But by taking a ratio,  $q_0$  cancels out, so it is not needed for estimating the risk ratio.