Double Negative Control Inference in Test-Negative Design Studies of Vaccine Effectiveness

Kendrick Qijun Li[*]
Department of Biostatistics, University of Michigan;
Xu Shi
Department of Biostatistics, University of Michigan;
Wang Miao
Department of Probability and Statistics, Peking University;
Eric Tchetgen Tchetgen
Department of Statistics and Data Science, The Wharton School,
University of Pennsylvania

April 4, 2023

Abstract

The test-negative design (TND) has become a standard approach to evaluate vaccine effectiveness against the risk of acquiring infectious diseases in real-world settings, such as Influenza, Rotavirus, Dengue fever, and more recently COVID-19. In a TND study, individuals who experience symptoms and seek care are recruited and tested for the infectious disease which defines cases and controls. Despite TND's potential to reduce unobserved differences in healthcare seeking behavior (HSB) between vaccinated and unvaccinated subjects, it remains subject to various potential biases. First, residual confounding may remain due to unobserved HSB, occupation as healthcare worker, or previous infection history. Second, because selection into the TND sample is a common consequence of infection and HSB, collider stratification bias may exist when conditioning the analysis on tested samples, which further induces confounding by latent HSB. In this paper, we present a novel approach to identify and estimate vaccine effectiveness in the target population by carefully leveraging a pair of negative control exposure and outcome variables to account for potential hidden bias in TND studies. We illustrate our proposed method with extensive simulations and an application to study COVID-19 vaccine effectiveness using data from the University of Michigan Health System.

Keywords: Causal inference, proximal causal inference, selection bias, unmeasured confounding

1

1 Introduction

1.1 Test-negative design studies of vaccine effectiveness

The test-negative design (TND) has become a standard approach to evaluate real-world vaccine effectiveness (VE) against the risk of acquiring infectious diseases (Sullivan et al. 2014, Chua et al. 2020). In an outpatient Influenza VE TND study, for example, symptomatic individuals seeking care and meeting eligibility criteria are enrolled and their Influenza virus infection status is subsequently confirmed via a laboratory test. VE against flu infection is then measured by comparing the prevalence of vaccination between the test-positive "cases" and test-negative "controls" (Jackson & Nelson 2013, Jackson et al. 2017). Besides Influenza, the TND and its variants have also been used to study VE against pneumococcal disease (Broome et al. 1980), dengue (Anders et al. 2018), rotavirus (Boom et al. 2010), and other infectious diseases. Recently, the TND has increasingly been used in post-licensure evaluation of COVID-19 VE (Patel et al. 2020, Dean et al. 2021, Hitchings et al. 2021, Thompson et al. 2021, Dagan et al. 2021, Olson et al. 2022).

Test-negative designs are believed to reduce unmeasured confounding bias due to healthcare-seeking behavior (HSB), whereby care seekers are more likely to be vaccinated, have healthier behaviors that reduce the risk of infection, and get tested when ill (Jackson et al. 2006). By restricting analysis to care seekers who are tested for the infection in view (e.g. Influenza or COVID-19), the vaccinated and unvaccinated are more likely to share similar HSB and underlying health characteristics. Misclassification of infection status is also reduced because the analysis is restricted to tested individuals (Jackson & Nelson 2013).

Sullivan et al. (2016) used directed acyclic graphs (DAG) to illustrate the rationale behind TND in the context of evaluating VE against flu infection, as shown in Figures 1(a) and (b). We denote flu vaccination status by A and flu infection by Y, so that the arrow $A \rightarrow Y$ represents VE against flu infection. Selection into the TND study sample, denoted

by S, is triggered by a subject experiencing flu-like symptoms, seeking care at clinics or hospitals, and getting tested for Influenza infection, hence the $Y \to S$ edge. Healthcare-seeking behavior, denoted by HSB, may affect S, A, and Y because subjects with certain healthcare-seeking proclivities may be more likely to seek care, take annual flu shots, and participate in healthy and preventive behaviors. The above variables may be subject to effects of other baseline covariates, such as age, season, and high-risk conditions, included in Figure 1 as confounders X. The TND presumes that by restricting recruitment to care seekers, the study subjects essentially have identical HSB; in other words, conditioning the analysis on $S = 1$ is equivalent to conditioning on HSB$= 1$, which would then completely control for HSB (Figure 1(b)). Measured covariates X are further adjusted for by including these factors in a logistic regression model or by inverse probability weighting (Bond et al. 2016, Thompson et al. 2021).

However, the TND remains subject to potential hidden bias. First, it is unrealistic that all study subjects seeking care are lumped into a single category HSB$= 1$. It may be more realistic that HSB is not a deterministic function of S and remains a source of confounding bias even after conditioning on S. Furthermore, there might be other mismeasured or unmeasured confounders, denoted by U. For example, healthcare workers are at increased risk of flu infection due to higher exposure to flu patients and are more likely to seek care and receive vaccination due to health agency guidelines (Black et al. 2018). Prior flu infection history may also be a source of confounding if it alters the likelihood of vaccination and care-seeking, while also providing immunity against circulating strains (Sullivan et al. 2016, Krammer 2019). These potential sources of confounding, if not properly accounted for, can result in additional confounding bias, as illustrated in Figure 1(c). Finally, collider stratification bias is likely present due to conditioning on S, which is a common consequence of HSB, other risk factors (X, U), and Influenza infection Y (Lipsitch et al. 2016). That is, conditioning on S unblocks the backdoor path $A \leftarrow (X, U, HSB) \to S \leftarrow Y$, which

would in principle be blocked if study subjects had identical levels of HSB and other risk factors (Sullivan et al. 2016).

Accounting for these potential sources of bias is well known to be challenging, and potentially infeasible without additional assumptions or data. This can be seen in Figure 1(d), which is a simplified version of Figure 1(c) where the unmeasured confounders U include individuals' occupation as a healthcare worker, previous flu infection, HSB, and so on. Figure 1(d) indicates that the unmeasured confounders U induce both confounding bias through the path $A \leftarrow U \rightarrow Y$ and collider stratification bias through the path $A \leftarrow U \rightarrow S \leftarrow Y$. In presence of both unmeasured confounding and collider bias, causal bounds may be available (Gabriel et al. 2022) but likely too wide to be informative; causal identification in TND therefore remains to date an important open problem in the causal inference literature which we aim to resolve.

1.2 Negative control methods

In recent years, negative control variables have emerged as powerful tools to detect, reduce, and potentially correct for unmeasured confounding bias (Lipsitch et al. 2010, Miao et al. 2018, Shi et al. 2020). The framework requires that at least one of two types of negative control variables are available which are *a priori* known to satisfy certain conditions: a negative control exposure (NCE) known to have no direct effect on the primary outcome; or a negative control outcome (NCO), known not to be an effect of the primary exposure. Such negative control variables are only valid and therefore useful to address unmeasured confounding in a given setting to the extent that they are subject to the same source of confounding as the exposure-outcome relationship of primary interest. Thus, an observed association between a valid NCE and the outcome (conditional on the exposure and observed covariates) or one between a valid NCO and the exposure can indicate the presence of residual confounding bias. For example, in a cohort study to investigate flu
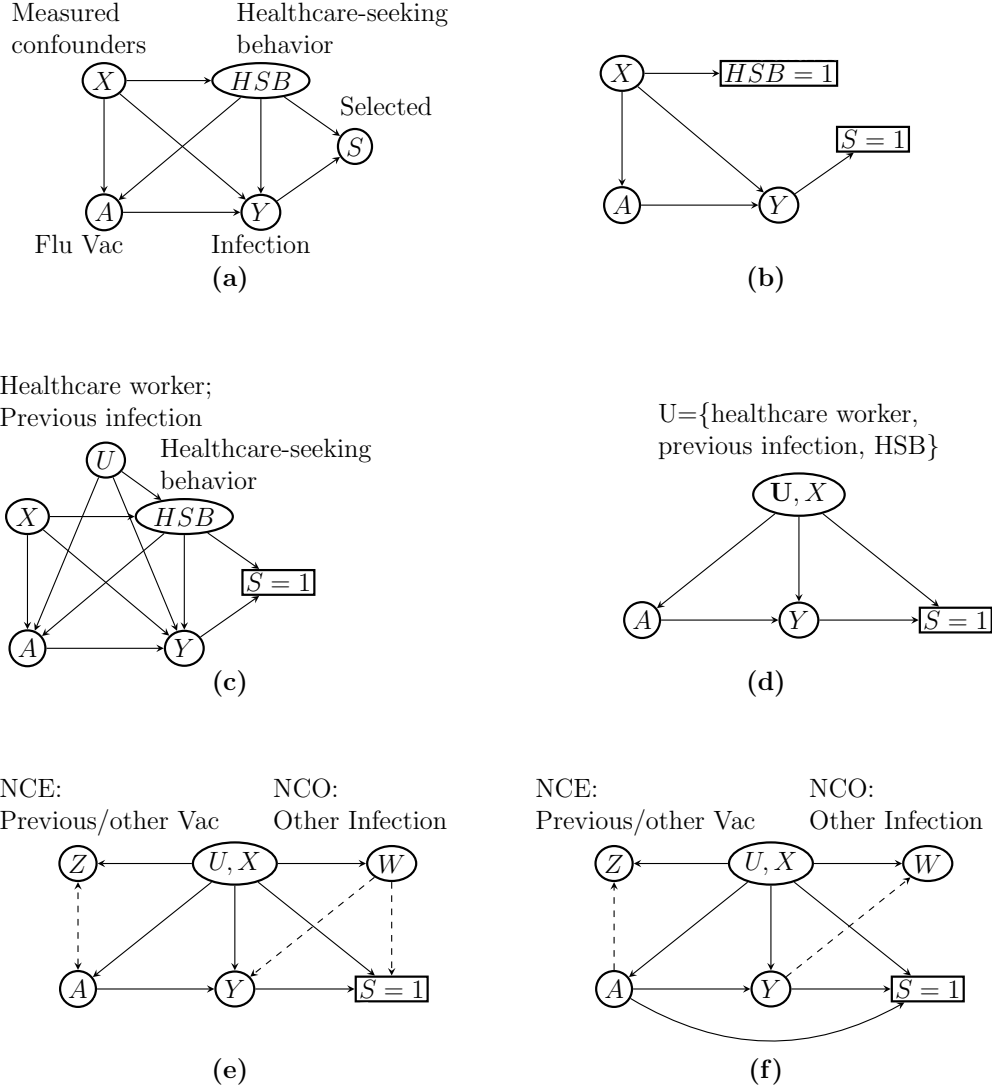
Figure 1: Causal relationships of variables in a test-negative design. Sullivan et al. (2016) used (a) to illustrate the causal relationship between variables in a test-negative design in the general population, and used (b) to illustrate the assumption implicit in the common approach to estimate VE from the study data that study subjects have identical healthcare-seeking behavior (HSB) (Sullivan et al. 2016). (c) shows that if HSB remains partially unobserved, then the backdoor paths $A \leftarrow HSB \rightarrow Y$ and $A \leftarrow HSB \rightarrow S = 1 \leftarrow Y$ indicate unmeasured confounding bias and selection bias, respectively. Other unmeasured confounders, such as occupation as a healthcare worker and previous infection, open additional backdoor paths between A and Y and result in additional confounding bias. (d) shows a simplified DAG from (c) that combines the unmeasured confounders into a single variable U. (e) illustrates our approach to estimate VE leveraging negative control exposure Z and outcome W. Dashed arrows indicate effects that are not required. (f) shows a scenario with the $A \rightarrow S$ arrow where the causal odds ratio can still be identified under additional assumptions.

VE against hospitalization and death among seniors, to detect potential unmeasured confounding, Jackson et al. (2006) used hospitalization/death before and after the flu season as NCOs and found that the association between flu vaccination and hospitalization was virtually the same before and during the flu season, suggesting that the lower hospitalization rate observed among vaccinated seniors versus unvaccinated seniors was partially due to healthy-user bias.

Recently, new causal methods have been proposed to not only detect residual confounding when present, but also to potentially de-bias an observational estimate of a treatment causal effect in the presence of unmeasured confounders when both an NCE and an NCO are available, referred to as the double negative control (Miao et al. 2018, Tchetgen Tchetgen et al. 2020). In this recent body of work, the double negative control design was extended in several important directions including settings in which confounding proxies routinely measured in well-designed observational studies may be used as negative control variables, a framework termed *proximal causal inference*; longitudinal settings where one is interested in the joint effects of time-varying exposures (Ying et al. 2023), potentially subject to both measured and unmeasured confounding by time-varying factors; and in settings where one aims to estimate direct and indirect effects in mediation analysis subject to unmeasured confounding or unmeasured mediators (Dukes et al. 2023, Ghassami et al. 2023). Other recent papers in this fast-growing literature include ?, Egami & Tchetgen Tchetgen (2023), Kallus et al. (2022), Imbens et al. (2021), Deaner (2021), Ghassami, Ying, Shpitser & Tchetgen Tchetgen (2022), and Ghassami, Shpitser & Tchetgen Tchetgen (2022). Notably, existing identification results in negative control and proximal causal inference literature have been restricted to i.i.d settings (Miao et al. 2020) and time series settings (Shi et al. 2023), and to date, to the best of our knowledge, outcome-dependent sampling settings such as TND have not been considered, particularly one where confounding and selection bias might co-exist.

1.3 Contribution and outline

In this paper, we introduced a novel double negative control approach to debias VE estimates from a test-negative design study. To our knowledge, this is the first work in the negative control method literature to address coexistent confounding and selection bias in the setting of a post-market vaccine effectiveness study. In Sections 2.1, we introduce notation and the identification under no unmeasured confounding nor selection bias . Next we develop our identification strategy and describe a new debiased estimator under a double negative control TND study in Sections 2.2-2.4, assuming no direct effect of vaccination on selection into the TND sample. In Section 2.5, we relax this assumption and introduce sufficient conditions under which VE remains identified. In Section 3, we demonstrate the performance of our method with simulations. In Section 4, the approach is further illustrated in an application to estimate COVID-19 VE against infection in a TND study nested within electronic health records from University of Michigan Health System. We conclude with a discussion in Section 5. We relegate all proofs, derivations, additional tables and figures, and detailed discussions to Section A-L of the Supplementary Materials.

2 Method

2.1 Preliminary: estimation under no unmeasured confounding and no selection bias

To fix ideas, we first review estimation assuming all confounders (U, X) are fully observed and the study sample is randomly drawn (rather than selected by testing) from a source population, referred to as the "target population". That is, we observe data on (A, Y, U, X) which are independent and identically distributed in the target population. For each individual, we write $Y(a)$ as the binary potential infection outcome had, possibly contrary to fact, the person's vaccination status been $A = a$, $a = 0, 1$.

Our goal is to provide identification and estimation strategies for the causal risk ratio (RR) defined as $RR = E[Y(1)]/E[Y(0)]$. Let β_0 denote the log causal RR, i.e., $RR = \exp(\beta_0)$. Following Hudgens & Halloran (2006), we define VE as one minus the causal RR: $VE = 1 - \exp(\beta_0)$. Let $Q(A = a, U, X) = 1/P(A = a|U, X)$ denote the inverse of the probability of vaccination status $A = a$ given confounders. Under the standard assumptions of consistency (which involves the assumption of no interference: a subject's potential outcome is not affected by the treatment of other subjects (Cole & Frangakis 2009)), ignorability (given U, X) and positivity (Hernán & Robins 2020), it is well known that, if U were observed, the mean potential outcome $E[Y(a)]$ can be identified by inverse probability of treatment weighting (IPTW):

$$P[Y(a) = 1] = E\left[I(A = a)Q(A = a, U, X)Y\right], \tag{1}$$

for $a = 0, 1$. Therefore, the log causal RR β_0 satisfies the following equation

$$E\left[Q(A = 1, U, X)AY \exp(-\beta_0)\right] - E\left[Q(A = 0, U, X)(1 - A)Y\right] = 0.$$

Equivalently, we have

$$E[V_0(A, Y, U, X; \beta_0)] = 0 \tag{2}$$

where $V_0(A, Y, U, X; \beta) = (-1)^{1-A}Q(A, U, X)Y \exp(-\beta A)$ is an unbiased estimating function for β_0.

2.2 Tackling selection bias under a semiparametric risk model

Next, consider a TND study for which data (A, Y, X, U) is observed only for the tested individuals with $S = 1$. Because S is influenced by other factors such as infection, the estimating function $V_0(A, Y, U, X; \beta_0)$ may not be unbiased with respect to the study sample; i.e. $E[V_0(A, Y, U, X; \beta_0) \mid U, X, S = 1] \neq 0$ without another assumption about the selection process into the TND sample.

For a TND sample of size n, we denote the i-th study subject's variables as (A_i, Y_i, U_i, X_i),

$i = 1, \ldots, n$. For generalizability, we make the key assumption that vaccination A is unrelated to selection S other than through a subject's infection status Y and confounders (U, X).

Assumption 1 (Treatment-independent sampling). $S \perp\!\!\!\perp A \mid Y, U, X$.

In a TND study, this assumption requires that an individual's decision to seek care and get tested only depends on the presence of symptoms and his/her underlying behavioral or socioeconomic characteristics, including HSB (contained in (U, X)), and therefore vaccination status does not directly affect selection. Such assumption is commonly made in previous works on test-negative design studies (Jackson & Nelson 2013, Sullivan et al. 2016) and selection bias (Didelez et al. 2010, Bareinboim & Pearl 2012), although the latent factor U is typically assumed absent in such literature. The DAGs in Figure 1(a)-(e) in fact encode this conditional independence condition. We relax this assumption in Section 2.5. Outcome-dependent sampling distinguishes our work from standard proximal causal inference, which has exclusively assumed a random sample of subjects from the target population (Miao et al. 2018, Cui et al. 2023, Tchetgen Tchetgen et al. 2020). We further consider the following effect homogeneity condition.

Assumption 2 (No effect modification). *For $a = 0, 1$,*

$$P(Y = 1|A = a, U, X) = \exp(\beta_0 a)g(U, X) \tag{3}$$

where $g(U, X) = P(Y = 1 \mid A = 0, U, X)$ is an unknown function only restricted by $0 \leq P(Y = 1 \mid A, U, X) \leq 1$.

Assumption 2 defines a semiparametric multiplicative risk model which states that vaccine effectiveness, measured on the RR scale, is constant across (U, X) strata in the target population. In other words, VE is not modified by U, X. This assumption is stronger than necessary for our methods but simplifies the exposition. In Section H of the Supplementary Materials, we relax the assumption to allow for effect modification by measured

confounders X. Infection risk for control subjects $P(Y = 1|A = 0, U, X) = g(U, X)$ is the nonparametric component of the model which is left unspecified.

Under Assumption 2, one can verify that $\exp(\beta_0) = E[Y(1)]/E[Y(0)]$, which is the marginal causal RR. Therefore, the estimating equation (2) implies that it is possible to identify β_0 even though the potential outcome means $E[Y(0)]$ and $E[Y(1)]$ cannot be identified due to selection bias. The following proposition indicates that the same is true when the data are subject to selection bias of a certain structure.

Proposition 1. *Under Assumptions 1 and 2, the parameter β_0 satisfies*

$$E[V_0(A, Y, U, X; \beta_0) \mid U, X, S = 1] = 0. \tag{4}$$

From Proposition 1, the IPTW estimating function V_0 derived from the target population remains unbiased in the TND sample. In principle, one could estimate β_0 with $\hat{\beta}$, the solution to

$$\frac{1}{n} \sum_{i=1}^{n} (-1)^{1-A_i} c(X_i) \widehat{Q}(A_i, U_i, X_i) Y_i \exp(-\hat{\beta} A_i) = 0, \tag{5}$$

where $c(\cdot)$ is a user specified function, and $\widehat{Q}(A_i, U_i, X_i) = 1/\widehat{P}(A = A_i \mid U_i, X_i)$ is the estimated probability of having vaccination status $A = A_i$ given confounders (U_i, X_i). Letting $c(X_i) = 1$, the resulting estimator

$$\hat{\beta}_0 = \log \left\{ \left[\sum_{i=1}^{n} \hat{Q}(A_i, U_i, X_i) A_i Y_i \right] / \left[\sum_{i=1}^{n} \hat{Q}(A_i, U_i, X_i)(1 - A_i) Y_i \right] \right\} \tag{6}$$

is essentially the IPTW estimator of marginal RR in Schnitzer (2022) assuming (U_i, X_i)'s are all observed.

However, $Q(A, U, X)$ cannot be estimated because U is unobserved. Furthermore, even if U were observed, $\widehat{Q}(A_i, U_i, X_i)$ may not be identified from the TND sample due to selection bias. In the next section, we describe a new framework to account for unmeasured confounding in a TND setting, leveraging negative control exposure and outcome variables.

2.3 Tackling unmeasured confounding bias leveraging negative controls

2.3.1 Negative control exposure (NCE) and the treatment confounding bridge function

As shown in Figure 1(e), suppose that one has observed a valid possibly vector-valued NCE, denoted as Z, which is a priori known to satisfy the following key independence conditions:

Assumption 3 (NCE independence conditions). $Z \perp\!\!\!\perp (Y, S)|A, U, X$.

Assumption 3 essentially states that any existing $Z - Y$ association conditional on (X, A) in the target population must be a consequence of their respective association with U, therefore indicating the presence of confounding bias. Importantly, the NCE must *a priori* be known to have no causal effect on the infection status (Miao et al. 2020). Likewise, the association between Z and S conditional on (X, A) is completely due to their respective association with U. Figure 1(e) presents a graphical illustration of an NCE that satisfies Assumption 3.

Shi et al. (2020) provided some general guidelines and examples of how to select an NCE in different settings. In the Influenza VE setting, a candidate NCE can be vaccination status for the preceding year, or other vaccination status such as Tdap (Tetanus, Diphtheria, Pertussis) vaccine, as both are known to effectively provide no protection against the circulating flu strain in a given year. We emphasize that an appropriate NCE should have no direct effect on selection into the TND sample. In other words, the selected NCE should be irrelevant to the study's inclusion/exclusion criteria other than through U, X. We now provide an intuitive description of our approach to leverage Z as an imperfect proxy of U for identification despite not directly observing U.

To motivate the rationale behind identification, ignore selection bias for now and suppose that $Q(A, U) = \alpha_0 + \alpha_1 A + \alpha_2 U$, also suppressing measured confounders X. Although

U is unobserved, suppose further that Z satisfies $E[Z|A, U] = \gamma_0 + \gamma_1 A + \gamma_2 U$. Then we have $U = E[\tilde{U}(A, Z)|A, U]$, where $\tilde{U}(A, Z) = (Z - \gamma_0 - \gamma_1 A)/\gamma_2$. Replacing U with $\tilde{U}(A, Z)$ in $Q(A, U)$, we get $q(A, Z) = \alpha_0 + \alpha_1 A + \alpha_2 \tilde{U}(A, Z)$, which does not depend on unmeasured confounder U and can recover the inverse probability of vaccination from $Q(A, U) = E[q(A, Z)|A, U]$. If all parameters of q were known, it would naturally follow that the IPTW method in Equation (1) can be recovered by

$$E[Y(a)] = E\{I(A = a)E[q(A, Z) \mid A, U]Y\} \overset{A.3}{=} E\left[I(A = a)q(A, Z)Y\right],$$

Therefore, β_0 can be identified if the distribution of (A, Y, Z) in the target population is available provided that parameters indexing q can be identified. The above insight motivates the following assumption:

Assumption 4 (treatment confounding bridge function). *There exists a function $q(A, Z, X)$ that satisfies, for every a, u and x,*

$$Q(A = a, U = u, X = x) = E\left[q(A, Z, X) \mid A = a, U = u, X = x\right] \tag{7}$$

A function q that satisfies (7) is called a treatment confounding bridge function (Cui et al. 2023). Below we give two examples where Equation (7) admits a closed form solution.

Example 1. *(Binary U and Z) Suppose that U is binary, and so is the NCE Z. For simplicity we suppress X. Write $p_{za.u} = P(Z = z, A = a \mid U = u)$. We prove in Section C of the Supplementary Materials that $q(a, z)$ has a closed form given by*

$$q(a, z) = \left[p_{1a.1} - p_{1a.0} + (p_{0a.0} - p_{0a.1} - p_{1a.1} + p_{1a.0})z\right] / (p_{0a.0}p_{1a.1} - p_{0a.1}p_{1a.0}). \tag{8}$$

The result can be similarly extended to polytomous Z.

Example 2. *(Continuous U and Z) Suppose the unmeasured confounder U and the NCE*

Z are both continuous. Further assume that

$$A|U, X \sim Bernoulli([1 + \exp(-\mu_{0A} - \mu_{UA}U - \mu_{XA}X)]^{-1})$$

$$Z|A, U, X \sim N(\mu_{0Z} + \mu_{AZ}A + \mu_{UZ}U + \mu_{XZ}X, \sigma_Z^2).$$

The treatment confounding bridge function $q(A, Z, X)$ can then be shown equal to

$$q(A, Z, X) = 1 + \exp\left[(-1)^A(\tau_0 + \tau_1 A + \tau_2 Z + \tau_3 X)\right] \tag{9}$$

where $\tau_0 = \mu_{0A} - \dfrac{\mu_{UA}\mu_{0Z}}{\mu_{UZ}} - \dfrac{\sigma_z^2 \mu_{UA}^2}{2\mu_{UZ}^2}$, $\tau_1 = \dfrac{\sigma_z^2 \mu_{UA}^2}{\mu_{UZ}^2} - \dfrac{\mu_{UA}\mu_{AZ}}{\mu_{UZ}}$, $\tau_2 = \mu_{UA}/\mu_{UZ}$, and $\tau_3 = \mu_{XA} - \mu_{XZ}\mu_{UA}/\mu_{UZ}$.

Formally, Equation (7) defines a Fredholm integral equation of the first kind, with $q(A, Z, X)$ as its solution (Cui et al. 2023). Heuristically, the existence of a solution requires that variation in Z induced by U is sufficiently correlated with variation in A induced by U. For instance, in Example 2, the existence of a treatment confounding bridge function amounts to the condition $\mu_{UZ} \neq 0$, which again requires $Z \not\perp U|A, X$. Cui et al. (2023) provided formal conditions sufficient for the existence of the treatment confounding bridge function satisfying Equation (7). These conditions are reproduced for completeness in Section B of the Supplementary Materials.

Thus, under Assumption 4, we propose to construct a new unbiased estimating function for β_0 by replacing $Q(A, U, X)$ with $q(A, Z, X)$ in $V_0(A, Y, U, X; \beta_0)$.

Theorem 1. *(Moment restriction for β_0) Under Assumptions 1-4, we have that*

$$E[V_1(A, Y, Z, X; \beta_0) \mid U, X, S = 1] = 0$$

where $V_1(A, Y, Z, X; \beta_0) = (-1)^{1-A}q(A, Z, X)Y \exp(-\beta_0 A)$.

Theorem 1 immediately implies that

$$E\left[(-1)^{1-A}c(X)q(A, Z, X)Y \exp(-\beta_0 A) \mid S = 1\right] = 0$$

for any function $c(X)$. In practice, if one can consistently estimate the treatment confounding bridge function $q(A, Z, X)$ with $\widehat{q}(A, Z, X)$, then β_0 can be estimated by solving

the estimating equation

$$\frac{1}{n}\sum_{i=1}^{n}(-1)^{1-A_i}c(X_i)\widehat{q}(A_i,Z_i,X_i)Y_i\exp(-\beta_0 A_i)=0,\tag{10}$$

for unidimensional $c(\cdot)\neq 0$, which results in the closed form estimator

$$\widehat{\beta}_0=\log\left(\frac{\sum c(X_i)\widehat{q}(A_i,Z_i,X_i)A_iY_i}{\sum c(X_i)\widehat{q}(A_i,Z_i,X_i)(1-A_i)Y_i}\right).$$

Importantly, although Equation (7) may not have a unique solution, any solution uniquely identifies the causal log RR β_0 for a fixed function $c(\cdot)$. Furthermore, although the choice of $c(\cdot)$ does not impair unbiasedness of (10), it does impact efficiency of the resulting estimator $\hat{\beta}_0$. In practice, one may simply set $c(X_i)\equiv 1$.

There remains the question of identifying and estimating the treatment confounding bridge function from the TND sample which we consider next.

2.3.2 Negative control outcome (NCO) for identification of treatment confounding bridge function

For identification and estimation of q, we propose to leverage NCOs to construct feasible estimating equations for the treatment confounding bridge function as in Cui et al. (2023). Similar to NCEs, NCOs can be viewed as imperfect proxies of U. However, unlike NCEs, a valid NCO, denoted by W, is (i) known a priori not to be a causal effect of either the primary exposure A or the NCE Z; and (ii) is associated with (A,Z) conditional on X only to the extent that it is associated with U. Formally, we make the following assumption.

Assumption 5. *(NCO Independence Conditions).*
(a) $W\perp\!\!\!\perp A\mid U,X$;   *(b)* $W\perp\!\!\!\perp Z\mid A,U,X,Y$;   *(c)* $S\perp\!\!\!\perp(A,Z)\mid U,X,W,Y$.

Assumptions 5(a) and (b) formalize the requirement that neither the exposure nor NCE can have direct effects on the NCO. Assumption 5(c) complements Assumption 3 and states that conditioning on W in addition to (A,U,X,Y) does not alter the conditional independence of Z with S. General guidelines for selecting NCOs are discussed by Shi et al.

(2020). In TND studies in which there may be multiple possible test-negative illnesses, an NCO may be selected as one or more of the test-negative illnesses. In flu VE studies, a candidate NCO can be an infection whose risk is not causally affected by either A or Z. For example, if the selected NCE is Tdap vaccination, then a potential NCO may be current-year respiratory syncytial virus infection, as its risk is unlikely to be affected by Influenza or Tdap vaccination. Recent outpatient visits for other acute illnesses can also serve as NCO, such as blepharitis, wrist/hand sprain, lipoma, ingrowing nail, etc. (Leung et al. 2011). In contrast with an NCE, an NCO can have a direct effect on the selection S. Figure 1(e) illustrates an NCO W that satisfies Assumptions 5(a) and (b).

Similar to Cui et al. (2023), we leverage an NCO as an additional proxy to identify the treatment confounding bridge function. However, a complication arises due to the lack of a random sample from the target population, a key requirement in the approach outlined in Cui et al. (2023). In general, it is not possible to obtain sufficient information about either the distribution of W or that of U in the target population from only the TND data without an additional structural assumption (Bareinboim & Pearl 2012). In the following, we avoid imposing such an additional structural assumption by leveraging an important feature of several infectious diseases such as Influenza and COVID-19; mainly that contracting such an infection at any point in time is a rare event in most target populations of interest, and therefore information from the target population relevant to estimating the treatment confounding bridge function can be recovered from the test-negative control group. Formally, we make the following rare disease assumption.

Assumption 6 (Rare infection). *There exist a small positive number $\delta > 0$ such that*

$$P(Y = 1 | A = a, W = w, U = u, X = x) \leq \delta, \qquad \textit{for almost every } a, w, u, x \qquad (11)$$

Assumption 6 states that infected subjects, whether vaccinated or not and regardless of their NCO, only constitute a small proportion of each (U, X) stratum in the target

population. This assumption implies that $\frac{1}{1-\delta} \leq \frac{P(A,Z|U=u,X=x,Y=0)}{P(A,Z|U=u,X=x)} \leq 1 - \delta$. Thus, under Assumptions 1, 3 and 6, $P(A = a, Z = z \mid U = u, X = x) \approx P(A = a, Z = z \mid U = u, X = x, Y = 0, S = 1)$ for all a, z, x, u. We now introduce a key property of the treatment confounding bridge function in Theorem 2.

Theorem 2 (Identification of the treatment confounding bridge function). *Under Assumptions 1, 3, 4, 5, and 6, for $a = 0, 1$ we have that*

$$\frac{(1-\delta)^3}{P(A = a \mid W, X, Y = 0, S = 1)} < E[q(a, Z, X) \mid W, A = a, X, Y = 0, S = 1]$$
$$< \frac{1}{(1-\delta)^3 P(A = a \mid W, X, Y = 0, S = 1)}$$

Thus, provided δ is small, Theorem 2 suggests that an approximation to the treatment confounding bridge function can be obtained by solving the following integral equation involving only observed data

$$E[q^*(A, Z, X) \mid W, A = a, X, Y = 0, S = 1] = 1/P(A = a \mid W, X, Y = 0, S = 1) \quad (12)$$

provided a solution exists. Accordingly, hereafter suppose that the following assumption holds.

Assumption 7 (Existence of a unique solution to (12)). *There exists a unique square-integrable function $q^*(A, Z, X)$ that satisfies (12).*

Heuristically, uniqueness of a solution to (12) requires that variation in W is sufficiently informative about variation in Z, in the sense that there is no variation in W that is not associated with corresponding variation in Z. See Section D of the Supplementary Materials for further elaboration of completeness condition and Newey & Powell (2003), D'Haultfoeuille (2011) for related use of the assumption in the literature. Below we briefly illustrate Assumption 7 in the examples of Section 2.3.1.

Example 1'. *Suppose U and Z are both binary, and a binary NCO W is also observed. Let $p'_{za.w} = P(Z = z, A = a \mid W = w, Y = 0, S = 1)$ for $z, a, w \in \{0, 1\}$, then solving Equa-*

tion (12) *gives* $q^*(a,z) = [p'_{1a.1} - p'_{1a.0} + (p'_{0a.0} - p'_{0a.1} - p'_{1a.1} + p'_{1a.0})z] / (p'_{0a.0}p'_{1a.1} - p'_{0a.1}p'_{1a.0})$. *The probabilities* $p'_{za.w}$ *can all be estimated from the study sample.*

We emphasize that the solution to Equation (12) is ultimately an approximation to the (non-identifiable) treatment confounding bridge function in the target population. The accuracy of this approximation relies on the extent to which the rare disease assumption holds in the target population of interest. We study the potential bias due to departure from this key assumption in Section E of the Supplementary Materials. We further observe that, under the null hypothesis of no vaccine effectiveness, or if W has no direct effects on Y or S, then the function $q^*(A, Z, X)$ matches the treatment confounding bridge exactly, even if the disease is not rare, as stated in the corollary below.

Corollary 1. *Under the Assumptions of Theorem 1, Assumption 7, and the null hypothesis of no vaccine effect against infection, such that* $Y \perp\!\!\!\perp A \mid U, X$, *then*

$$E[q(A, Z, X) \mid W, A = a, X, Y = 0, S = 1] = 1/P(A = a \mid W, X, Y = 0, S = 1).$$

From Theorem 2, we immediately have the following corollary which provides a basis for estimation of $q^*(A, Z, X)$ from the observed TND sample.

Corollary 2. *Under the conditions of Theorem 2, for any function* $m(W, A, X)$, *the solution* $q^*(A, Z, X)$ *to Equation* (12) *also solves the population moment equation*

$$E\left[m(W, A, X)q^*(A, Z, X) - m(W, 1, X) - m(W, 0, A) \mid Y = 0, S = 1\right] = 0. \qquad (13)$$

In practice, a parametric model $q^*(A, Z, X; \tau)$ for the treatment confounding bridge function might be appropriate, where τ is an unknown finite dimensional parameter, Corollary 2 suggests one can then estimate τ by solving the estimating equation

$$\frac{1}{n}\sum_{i=1}^{n}(1 - Y_i)[m(W, A, X)q(A, Z, X; \hat{\tau}) - m(W, 1, X) - m(W, 0, X)] = 0, \qquad (14)$$

where $m(W, A, X)$ is a user-specified function with dimension no smaller than τ's.

Example 1″. *If Z and W are both binary, rather than solving the system of equations implied by (12), one can instead specify a saturated model*

$$q^*(A, Z; \tau) = \tau_0 + \tau_1 Z + \tau_2 A + \tau_3 Z A \tag{15}$$

and estimate $\tau = (\tau_0, \tau_1, \tau_2, \tau_3)^T$ by solving Equation (14) with $m(W, A) = (1, W, A, WA)^T$. Extension to Z and X with multiple categories is relatively straightforward.

Example 2′. *In case of continuous (U, X, Z), Equation (9) suggests the model*

$$q^*(A, Z, X; \tau) = 1 + \exp\left[(-1)^A(\tau_0 + \tau_1 A + \tau_2 Z + \tau_3 X)\right]. \tag{16}$$

If a univariate NCO W is available, we may solve (14) with $m(W, A, X)$ defined as a vector including (W, A, X), their high-order interactions and an intercept.

2.4 Estimation and Inference

In previous sections, we have defined the structural parameter of interest β_0 as stratum-specific log risk ratio, introduced the treatment confounding bridge function as a key ingredient to identification of β_0, and presented a strategy to estimate the treatment confounding bridge function leveraging an NCO. We summarize the steps of our estimation framework in Algorithm 1 and present the large-sample properties of the resulting estimator $(\widehat{\beta}, \widehat{\tau})$ in Theorem 3.

Algorithm 1 Negative control method to estimate vaccine effectiveness from a test-negative design

---

1: Identify the variables in the data according to Figure 1(e), in particular the NCEs and NCOs.
2: Estimate the treatment confounding bridge function by solving Equation (14) with a suitable parametric model $q^*(A, Z, X; \tau)$ and a user-specified function $m(W, A, X)$. Write $\widehat{\tau}$ as the resulting estimate of τ.
3: Estimate β_0 by
$$\widehat{\beta} = \log\left(\frac{\sum c(X_i) q^*(a, Z_i, X_i; \widehat{\tau}) A_i Y_i}{\sum c(X_i) q^*(A_i, Z_i, X_i; \widehat{\tau})(1 - A_i) Y_i}\right)$$
where $\widehat{q}(A, Z, X) = q^*(A, Z, X; \widehat{\tau})$ and $c(\cdot)$ is a user-specified one-dimensional function.

Theorem 3 (Inference based on $(\widehat{\beta}, \widehat{\tau})$). *Under Assumptions 1-7 and suitable regularity conditions provided in Section **??**, the estimator $(\widehat{\beta}, \widehat{\tau})$ in Algorithm 1, or equivalently, the solution to the estimating equation $\frac{1}{n} \sum_{i=1}^n G_i(\beta, \tau) = 0$ is regular and asymptotically linear with the i-th influence function $IF_i(\beta, \tau) = -\Omega(\beta, \tau)^+ G_i(\beta, \tau)$, where $\Omega(\beta, \tau)^+$ denotes the Moore-Penrose inverse of $\Omega(\beta, \tau)$,*

$$G_i(\beta, \tau) = \begin{pmatrix} (-1)^{1-A_i} c(X_i) q^*(A_i, Z_i, X_i; \tau) Y_i \exp(-\beta A_i) \\ (1 - Y_i) \left[m(W_i, A_i, X_i) q^*(A_i, Z_i, X_i; \tau) - m(W_i, 1, X_i) - m(W_i, 0, X_i) \right] \end{pmatrix}$$

and $\Omega(\beta, \tau) = \left(E\left[\partial G_i(\beta, \tau)/\partial \beta^T \right], E\left[\partial G_i(\beta, \tau)/\partial \tau^T \right] \right).$

Large-sample standard errors and confidence intervals follow from standard estimating equation theory (See Van der Vaart (2000) Theorem 5.21).

The estimator $\hat{\beta}$ and its standard error are constructed under Assumption 6. If Assumption 6 fails to hold, $\hat{\beta}$ may be biased and confidence intervals may not be well-calibrated. However, by Corollary 1, under the null hypothesis of no vaccine effect, the estimated $q^*(A, Z, X)$ converges to the true treatment confounding bridge function and $\hat{\beta}$ is consistent for $\beta_0 = 0$. This implies that while our methods are approximately asymptotically unbiased for rare infections, they provide a valid test of no vaccine causal effect even if the infection is not rare.

2.5 Estimating VE under treatment-induced selection

Thus far, unbiasedness of the estimating function V_0 has crucially relied on Assumption 1 that A does not have a direct effect on S. In some settings, this assumption may be violated if an infected person who is vaccinated is on average more likely to present to the ER than an unvaccinated infected person with similar symptoms, so that treatment or vaccination-induced selection into the TND sample is said to be present. In such settings, the estimator $\hat{\beta}$ produced by Algorithm 1 may be severely biased. Crucially, we note that this form of selection bias can be present even in context of a randomized trial in which

19

vaccination/treatment is known to the analyst, if the outcome is ascertained using a TND, for example in recent cluster-randomized test-negative design studies of community-level dengue intervention effectiveness (Anders et al. 2018, Jewell et al. 2019, Dufault & Jewell 2020, Wang et al. 2023). In this section, we provide sufficient conditions for identification under treatment-induced selection. In this vein, consider the following assumptions:

Assumption 1'. $P(S = 1 \mid A = a, Y = 1, U, X)/P(S = 1 \mid A = a, Y = 0, U, X) = \exp(h(U, X))$ *for $a = 0, 1$.*

That is, the risk ratio association between infection status and selection into the TND sample is independent of vaccination status. Equivalently, the conditional probability of selection can be factorized as $P(S = 1 \mid A, Y, U, X) = h_1(A, U, X)h_2(Y, U, X)$. This assumption may be reasonable if, given (U, X), the treatment A and outcome Y affect sample selection through independent mechanisms such that A does not modify the effect of Y on sample selection S on the multiplicative scale. Assumptions similar to Assumption 1' have been used previously to illustrate settings where two independent causes of sample selection remain conditionally independent in the selected sample (see Hernán et al. (2004) Appendix A.3). We highlight that Assumption 1 (which requires $A \perp\!\!\!\perp S \mid U, X, Y$) implies Assumption 1' and therefore is more stringent than the latter assumption. Furthermore,

Assumption 2'. *(No effect modification by confounders on the OR scale).*
$$\frac{P(Y = 1 \mid A = 1, U, X)/P(Y = 0 \mid A = 1, U, X)}{P(Y = 1 \mid A = 0, U, X)/P(Y = 0 \mid A = 0, U, X)} = \exp(\beta'_0).$$

Recall that Assumption 2 posited a constant vaccination causal effect on the RR scale across levels of (U, X), while Assumption 2' instead posits that the corresponding causal effect on the odds ratio scale is constant. In case of a rare infection in the target population, the OR and RR are approximately equal, in which case VE is well approximated by $1 - OR$.

Furthermore, identification relies on the following modified definition of a treatment confounding bridge function:

Assumption 4′. *There exists a treatment confounding bridge function $\tilde{q}$ such that for $a = 0, 1$,*

$$E[\tilde{q}(a, Z, X) \mid A = a, U, X] = 1/P(A = a \mid U, X, Y = 0, S = 1) \qquad almost\ surely. \quad (17)$$

We now obtain the identification of the OR in the following theorem:

Theorem 1′. *Under Assumptions 1′, 2′, 3 and 4′, we have that*

$$E[\tilde{V}_1(A, Y, Z, X; \beta_0') \mid U, X, S = 1] = 0$$

where $\tilde{V}_1(A, Y, Z, X; \beta)$ is defined as V_1 in Theorem 1 except with $\tilde{q}$ replacing q.

Importantly, the theorem establishes that the estimating function V_1 previously developed in the paper can, under stated conditions, remain unbiased for the odds ratio association of vaccination with testing positive for the infection, even in the presence of treatment-induced selection into the TND sample.

Estimation of the treatment confounding bridge function $\tilde{q}(A, Z, X)$ requires a negative control outcome that satisfies:

Assumption 5′. *(NCO Independence Conditions)* $W \perp\!\!\!\perp (A, Z, S) \mid U, X, Y$.

In addition to Assumption 5, this last assumption requires that neither Y nor S is a causal effect of W. Figure 1(f) illustrates a DAG that satisfies our assumptions regarding (Z, W). As can be verified in the graph, Assumption 5′ is needed to ensure that collider stratification bias induced by the path $A \to [S = 1] \leftarrow W$ upon conditioning on $S = 1$ is no longer present. Identification of the function $\tilde{q}$ is given below:

Theorem 2′. *Under Assumptions 3, 4′ and 5′, for $a = 0, 1$ we have that*

$$E[\tilde{q}(a, Z, X) \mid A = a, W, X, Y = 0, S = 1] = 1/P(A = a \mid W, X, Y = 0, S = 1)$$

As a result of Theorem 2′, the parameters in the treatment confounding bridge function can be estimated by solving Equation (14).

In summary, the above discussion suggests that one can continue to use Algorithm 1 to estimate VE in presence of treatment induced selection bias, albeit on the OR scale and under a modified set of negative control conditions. Theorem 3 continues to apply.

As a side note, Assumption 1′ automatically holds under Assumption 1, hence the above results also apply to the setting in previous sections illustrated in Figure 1(e). We state this result in the following corollary.

Corollary 3. *Under Assumptions 1, 2′, 3 and 4′, we have*

$$E[\tilde{V}_1(A, Y, Z, X; \beta'_0) \mid U, X, S = 1] = 0.$$

With Assumption 1, the treatment confounding bridge function $\tilde{q}$ can be estimated by solving tquation (14) either under Assumption 5′ and 7, or Assumptions 5, 6 and 7 as an approximation under the rare disease assumption. Corollary 3 leads to an interesting observation: under treatment-independent sampling (Assumption 1), the estimator $\hat{\beta}$ from Algorithm 1 can be viewed as either log RR or log OR, depending on the setting.

Finally, as Schnitzer (2022) pointed out, under the additional assumptions that (1) subjects included in the TND analysis who do not test positive for the infection in view are nevertheless subject to another infection, referred to as a test-negative infection, (2) test-positive and test-negative infections are mutually exclusive, and (3) vaccination of interest has no causal effect on the test-negative infection, the conditional odds ratio $\exp(\beta_0)$ also identifies the conditional causal risk ratio. We further discussed this result in Section G of the Supplementary Materials.

3 Simulation Study

To assess the empirical performance of our proposed method, we perform simulation studies mimicking a TND with binary vaccination, infection, and testing. We generate a target population of $N = 7,000,000$ according to Figure 1(e). We consider two scenarios where U,

Z and W are binary or continuous. In each scenario, we consider the true value of the log risk ratio $\beta_0 = \log(0.2), \log(0.5), \log(0.7)$ and 0. We set baseline log risk $\eta_0 = \log(0.01)$ so that the outcome is rare in the target population. The generated TND sample had sample size raging between $n = 43,000$ and $n = 52,000$.

For each scenario, we evaluated the performance of bias and coverage rates of 95% confidence intervals of the proposed NC estimator for β_0 over 500 Monte Carlo samples. As the estimated treatment confounding bridge function in Algorithm 1 is only an approximation under Assumption 6, whose bias may affect the estimation for β_0, we also include the NC-Oracle estimator that uses the true treatment confounding bridge function. For comparison, we included two standard estimators for β_0 in test-negative designs, logistic regression (Bond et al. 2016) and the IPTW estimator (Schnitzer 2022), both of which do not account for bias due to the latent factor U.

Due to the page limit, we relegated the details of the data generating process and figures for the results to Section I of the Supplementary Materials. Figure S.1 in Section I reports the bias of the four estimators we considered and coverage probability of their 95% confidence intervals. In both settings, both NC and NC-Oracle are essentially unbiased whereas logistic regression and IPTW give biased estimates in all scenarios. NC-Oracle exhibits slightly higher precision than NC, which implies that estimating the treatment confounding bridge function in the TND is only slightly more variable. Confidence intervals of NC and NC-Oracle both attain nominal coverage, whereas logistic regression and IPTW based confidence intervals undercover severely.

To investigate the performance of our method for a non-rare infection, we repeat the simulation under the same setup but increasing baseline log risk η_0. In our simulation setting, $\exp(\eta_0)$ equals δ in Assumption 6 that describes the prevalence of the infection outcome in the target population, which determines the bias of estimating the treatment confounding bridge function. The results are reported in Figures S.2 and S.3 in Section I

of the Supplementary Materials. While the NC-Oracle estimator remains unbiased and maintained calibrated confidence intervals, the bias of the NC estimator increases with increasing δ and has anti-conservative 95% CIs. With $\delta = 0.5$ and $\beta_0 = \log(0.2)$, the bias of the NC estimator increases to -0.27 (corresponding to 5% overestimation on the VE scale). The coverage of 95% CI for the NC estimator drops to 0% when $\delta \geq 0.2$. Notably, the NC estimator remains unbiased with calibrated confidence intervals under the null hypothesis $\beta_0 = 0$, which suggests a Wald test of no vaccine effect still has the correct size. Both logistic regression and IPTW estimators are severely biased in all settings.

4 Application

We applied our proposed method to a TND study of COVID-19 VE of two-dose Moderna vaccine (mRNA-1273), two-dose Pfizer-BioNTech vaccine (BNT162b2), and single-dose Johnson & Johnson's Janssen vaccine (Ad26.COV2.S) against SARS-Cov-2 infection nested in the University of Michigan Health System. The selected study sample included patients who interacted with the University of Michigan Health System and experienced COVID-19 symptoms, had suspected exposure to COVID-19 virus, or sought to screen for SARS-Cov-2 infection, between April 5, 2021 and December 7, 2021. In addition, selected test-positive subjects had at least one positive lab test for SARS-Cov-2 infection after April 5. Vaccination history was obtained through electronic health records. A study subject was considered fully vaccinated if they received at least one dose of Johnson & Johnson's Janssen vaccine or at least two doses of Moderna or Pfizer-BioNTech vaccine. If a subject tested positive before or within 14 days after their first dose of Janssen vaccine or within 14 days after their second dose of Moderna or Pfizer-BioNTech vaccine, they were considered unvaccinated (Moline et al. 2021).

We selected immunization visits before December 2020 for NCE since COVID-19 vaccines were unavailable before December 2020 and prior immunization was unlikely to affect

SARS-Cov-2 infection risk; nor that of the selected NCOs we describe next. For NCO, we selected a binary indicator of having at least one of the following "negative control outcome" conditions after April 5, 2021: arm/leg cellulitis, eye/ear disorder, gastro-esophageal disease, atopic dermatitis, injuries, and general adult examination visits. Such candidate NCE and NCO are likely to satisfy the requisite conditional independence conditions of valid negative control variables and to be related to a patient's latent HSB. We adjusted for age group (<18, between 18 and 60, or ≥ 60), gender, race (white or non-white), Charlson comorbidity score ≥ 3, and the calendar month of a test-positive subject's first positive COVID test or a test-negative subject's last COVID test. Table S.2 in Section J of the Supplementary Materials summarizes the distribution of negative control variables, demographic variables and SARS-Cov-2 infection among vaccinated and unvaccinated subjects.

Because NCE is expected not to be associated with either the outcome or NCO in a fully adjusted analysis unless there is unmeasured confounding, we first fit regression models to detect presence of residual confounding bias. Conditioning on the baseline covariates, in both vaccinated and unvaccinated groups, NCE was significantly associated with SARS-Cov-2 infection ($p < 0.001$) and NCO ($p < 0.001$) in corresponding adjusted logistic regression models, suggesting the presence of hidden bias (See Section J Table S.3 and S.4 of the Supplementary Materials).

We implemented Algorithm 1 to estimate VE. We specified a linear model for the treatment confounding bridge function with an interaction term between COVID-19 vaccination and the NCE, and set the function m to include an intercept term, COVID-19 vaccination, the NCO, and baseline covariates, as well as all two-way interactions. For comparison, we implemented adjusted logistic regression model and the IPTW estimator of marginal risk ratio of Schnitzer (2022), where propensity score is estimated with logistic regression.

Table 1 shows the estimated VE and 95% confidence intervals for the three estimators. The VE estimates with our double NC method are notably higher than those of standard

logistic regression and IPTW estimator for all three vaccines.

Table 1: Estimated VE and 95% confidence intervals of the negative control estimator, logistic regression and IPTW estimator with the University of Michigan Health System data.

	Negative control	Logistic regression	IPTW
Pfizer-BioNTech	80.2% (78.3%, 81.9%)	74.1% (72.3%, 75.8%)	67.9% (66.0%, 69.6%)
Moderna	89.7% (88.1%, 91.1%)	78.8% (76.8%, 80.7%)	72.4% (69.5%, 75.1%)
Janssen (J & J)	65.8% (54.6%, 74.1%)	56.3% (48.4%, 62.9%)	48.2% (28.3%, 62.6%)

Recent TND studies estimated the mRNA COVID-19 vaccine (Pfizer-BioNTech and Moderna) effectiveness ranging between 80% and 98% against lab-confirmed SARS-COV-2 infection of different variants (Dagan et al. 2021, Bruxvoort et al. 2021, Israel et al. 2021). Our NC approach provided VE estimates that are closer to these prior studies. We hypothesize that the standard logistic regression and IPTW estimator underestimate the VE by overlooking residual confounding bias due to HSB and related factors, which our proposed double NC approach appears to control to some extent.

5   Discussion

In this article, we have introduced a statistical approach leveraging negative control variables to account for hidden bias due to residual confounding and/or selection mechanism in a test-negative design, both of which have raised major concerns (Sullivan et al. 2016, Schnitzer 2022). Negative control variables abound in practice, such as vaccination history which is routinely collected in insurance claims and electronic health records. Hence the proposed method may be particularly useful in such real-world settings to obtain improved estimates of vaccine effectiveness. Beyond TND, our method is also applicable to other study designs with outcome-dependent sampling, such as a case-control study for a rare disease where unmeasured confounding is of concern. With simple modifications, our approach can also be applied to settings with polytomous treatment and/or outcome, as

discussed in Section K of the Supplementary Materials.

The TND is a challenging setting in causal inference where selection bias and unmeasured confounding co-exist, the selection is outcome-dependent, and unmeasured confounders also impact selection. Jackson et al. (2018) performed simulation studies to evaluate the selection bias of VE estimate by logistic regression and concluded that selection bias due to healthcare-seeking behavior is unlikely to be meaningful in practice. However, their simulation study did not consider the effect of healthcare-seeking behavior on the infection outcome. Didelez et al. (2010) used graphical models to study conditions for the estimability of the odds ratio and testability of the hypothesis of null causal effect under outcome-dependent sampling, but they did not consider the setting where the unmeasured confounders also affect selection. A widely adopted approach to adjust for selection bias is to reweight each observation's contribution by their corresponding inverse probability of selection into the sample (Hernán et al. 2004), but such weights are unlikely to be available in most TNDs without access to a random sample from the target population and accurate measurements of latent healthcare-seeking behavior. Bareinboim et al. (2014) formally showed that causal effects cannot be recovered from outcome-dependent sampling without an additional assumption. Bareinboim & Pearl (2012) also showed that the causal odds ratio cannot be recovered under both confounding and selection bias. We have established that, however, progress can be made under a semiparametric multiplicative model, provided the outcome is rare in the target population, and that double negative control variables are available. To this end, this article showcases the potential power of negative control methods and proximal causal inference in epidemiologic research (Shi et al. 2020, Tchetgen Tchetgen et al. 2020).

We primarily focused on outpatient TND studies, where recruitment is restricted to subjects who seek care voluntarily. TNDs have also been applied to inpatient settings for studying VE against, for example, flu hospitalization (Foppa et al. 2016, Feng et al. 2016).

In inpatient TNDs, differential access to healthcare and underlying health characteristics between vaccinated and unvaccinated subjects are likely the main source of confounding bias (Feng et al. 2016). Our methods are still applicable in such settings, but negative control variables should be selected to be relevant to this source of unmeasured confounding. For example, previous vaccination and hospitalization outside of the flu season or hospitalization due to other flu-like illness are viable candidate NCE and NCO, respectively (Jackson et al. 2006).

Our approach is also suitable for post-market TND studies where real-world VE is of interest and vaccination history is obtained retrospectively, possibly through electronic health records. For vaccine efficacy in a controlled trial setting, Wang et al. (2023) recently developed estimation and inference of RR in cluster-randomized TND, aiming to correct for bias due to intervention-induced confounding by HSB induced due to unblinding. They proposed a log-RR estimator which corrects for selection bias by leveraging a valid test-negative outcome, under an assumption that either (i) the vaccine does not have a causal effect in the population, and the causal impact of vaccination on selection is equal for test-positive and -negative subsamples; or (ii) among care seekers, the incidence of test-negative outcomes does not differ between vaccinated and unvaccinated, and the intervention effect among care seekers is generalizable to the whole population. Even under randomization, identification conditions given in Section 2.5 are neither stronger nor weaker than those of Wang et al. (2023) described above, as neither set of assumptions appear to imply the other. An important advantage of our proposed methods is that they can be used to account for selection bias in a TND study irrespective of randomization.

Our methods target RR as a measure of VE instead of the more common OR (Jackson et al. 2006, Sullivan et al. 2016). These two measures are approximately equal for rare infections as described in Section L of the Supplementary Materials. Schnitzer (2022) recently considered estimation of a marginal causal RR in the TND sample and justified

the use of an inverse probability of treatment weighted (IPTW) estimator in a setting in which an unmeasured common cause of infection and selection into the TND sample does not cause vaccination (and thus there is no unmeasured confounding). Instead, our methods allow for an unmeasured common cause of vaccination, infection, and selection into the TND sample; however, in order to estimate a causal RR, we invoke both, an assumption of no effect modification by an unmeasured confounder, and a rare-disease condition. As we establish, the latter assumption can be relaxed to test the null hypothesis of no causal effect of the vaccine on infection risk. In Section 2.5, we establish that under a homogeneous OR condition, and an alternative definition of the treatment bridge function, our methods can identify a causal effect of the vaccine on the OR scale without invoking the rare disease condition.

Throughout the article, we have assumed diagnostic tests are accurate and individuals who seek care are sparsely distributed, such that the vaccination of a given subject in the TND sample does not protect another study subject from infection, i.e. there is no interference in the TND sample, a common assumption in TND literature. This latter assumption may be violated if members of the same households are present in the ER, in which case block interference must be accounted for using results from interference literature (Hudgens & Halloran 2008, Tchetgen Tchetgen & VanderWeele 2012).

We have proposed a parametric approach to estimate treatment confounding bridge function. While we have provided examples where certain parametric models are appropriate, in general, such a parametric approach may result in model misspecification bias. Nonparametric methods that have been developed for proximal causal inference, such as kernel machine learning (Ghassami, Ying, Shpitser & Tchetgen Tchetgen 2022) may be adapted to our setting. Another potential nonparametric approach is sieve generalized method of moments that uses basis functions to approximate the nuisance function (Ai & Chen 2003, Chen 2007). We leave these topics for future research.

Acknowledgements

Supplementary Materials

The Supplementary Materials contain proofs and derivations of all technical results presented in the paper and further discussions. All code needed to run the experiments given in the paper is now available at `https://github.com/KenLi93/TND_methods/`.

References

Ai, C. & Chen, X. (2003), 'Efficient estimation of models with conditional moment restrictions containing unknown functions', *Econometrica* **71**(6), 1795–1843.

Anders, K. L., Indriani, C., Ahmad, R. A., Tantowijoyo, W., Arguni, E., Andari, B., Jewell, N. P., Rances, E., O'Neill, S. L., Simmons, C. P. et al. (2018), 'The awed trial (applying wolbachia to eliminate dengue) to assess the efficacy of wolbachia-infected mosquito deployments to reduce dengue incidence in yogyakarta, indonesia: study protocol for a cluster randomised controlled trial', *Trials* **19**(1), 1–16.

Bareinboim, E. & Pearl, J. (2012), 'Controlling selection bias in causal inference', *Artificial Intelligence and Statistics* pp. 100–108.

Bareinboim, E., Tian, J. & Pearl, J. (2014), 'Recovering from selection bias in causal and statistical inference', *Proceedings of the AAAI Conference on Artificial Intelligence* **28**(1).

Black, C. L., Yue, X., Ball, S. W., Fink, R. V., de Perio, M. A., Laney, A. S., Williams, W. W., Graitcer, S. B., Fiebelkorn, A. P., Lu, P.-J. et al. (2018), 'Influenza vaccination coverage among health care personnel—united states, 2017–18 influenza season', *Morbidity and Mortality Weekly Report* **67**(38), 1050.

Bond, H., Sullivan, S. & Cowling, B. (2016), 'Regression approaches in the test-negative study design for assessment of influenza vaccine effectiveness', *Epidemiology & Infection* **144**(8), 1601–1611.

Boom, J. A., Tate, J. E., Sahni, L. C., Rench, M. A., Hull, J. J., Gentsch, J. R., Patel, M. M., Baker, C. J. & Parashar, U. D. (2010), 'Effectiveness of pentavalent rotavirus vaccine in a large urban population in the united states', *Pediatrics* **125**(2), e199–e207.

Broome, C. V., Facklam, R. R. & Fraser, D. W. (1980), 'Pneumococcal disease after pneumococcal vaccination: an alternative method to estimate the efficacy of pneumococcal vaccine', *New England Journal of Medicine* **303**(10), 549–552.

Bruxvoort, K. J., Sy, L. S., Qian, L., Ackerson, B. K., Luo, Y., Lee, G. S., Tian, Y., Florea, A., Aragones, M., Tubert, J. E. et al. (2021), 'Effectiveness of mRNA-1273 against delta, mu, and other emerging variants of SARS-CoV-2: test negative case-control study', *British Medical Journal* **375**.

Chen, X. (2007), 'Large sample sieve estimation of semi-nonparametric models', *Handbook of Econometrics* **6**, 5549–5632.

Chua, H., Feng, S., Lewnard, J. A., Sullivan, S. G., Blyth, C. C., Lipsitch, M. & Cowling, B. J. (2020), 'The use of test-negative controls to monitor vaccine effectiveness: a systematic review of methodology', *Epidemiology* **31**(1), 43.

Cole, S. R. & Frangakis, C. E. (2009), 'The consistency statement in causal inference: a definition or an assumption?', *Epidemiology* **20**(1), 3–5.

Cui, Y., Pu, H., Shi, X., Miao, W. & Tchetgen Tchetgen, E. (2023), 'Semiparametric proximal causal inference', *Journal of the American Statistical Association* (in press).

Dagan, N., Barda, N., Kepten, E., Miron, O., Perchik, S., Katz, M. A., Hernán, M. A., Lipsitch, M., Reis, B. & Balicer, R. D. (2021), 'BNT162b2 mRNA Covid-19 vaccine in a nationwide mass vaccination setting', *New England Journal of Medicine* .

Dean, N. E., Hogan, J. W. & Schnitzer, M. E. (2021), 'Covid-19 vaccine effectiveness and the test-negative design', *New England Journal of Medicine* **385**(15), 1431–1433.

Deaner, B. (2021), 'Proxy controls and panel data', *arXiv:1810.00283v7 [econ.EM]* .

Didelez, V., Kreiner, S. & Keiding, N. (2010), 'Graphical models for inference under outcome-dependent sampling', *Statistical Science* **25**(3), 368–387.

Dufault, S. M. & Jewell, N. P. (2020), 'Analysis of counts for cluster randomized trials: Negative controls and test-negative designs', *Statistics in Medicine* **39**(10), 1429–1439.

Dukes, O., Shpitser, I. & Tchetgen Tchetgen, E. (2023), 'Proximal mediation analysis', *Biometrika* (in press).

D'Haultfoeuille, X. (2011), 'On the completeness condition in nonparametric instrumental problems', *Econometric Theory* **27**(3), 460–471.

Egami, N. & Tchetgen Tchetgen, E. (2023), 'Identification and estimation of causal peer effects using double negative controls for unmeasured network confounding', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* (in press).

Feng, S., Cowling, B. J. & Sullivan, S. G. (2016), 'Influenza vaccine effectiveness by test-negative design–comparison of inpatient and outpatient settings', *Vaccine* **34**(14), 1672–1679.

Foppa, I. M., Ferdinands, J. M., Chaves, S. S., Haber, M. J., Reynolds, S. B., Flannery, B. & Fry, A. M. (2016), 'The case test-negative design for studies of the effectiveness of influenza vaccine in inpatient settings', *International Journal of Epidemiology* **45**(6), 2052–2059.

Gabriel, E. E., Sachs, M. C. & Sjölander, A. (2022), 'Causal bounds for outcome-dependent sampling in observational studies', *Journal of the American Statistical Association* **117**(538), 939–950.

Ghassami, A., Shpitser, I. & Tchetgen Tchetgen, E. (2022), 'Combining experimental and observational data for identification of long-term causal effects', *arXiv:2201.10743v3 [stat.ME]* .

Ghassami, A., Shpitser, I. & Tchetgen Tchetgen, E. (2023), 'Causal inference with hidden mediators', *arXiv:2111.02927v2 [stat.ST]* .

Ghassami, A., Ying, A., Shpitser, I. & Tchetgen Tchetgen, E. (2022), 'Minimax kernel machine learning for a class of doubly robust functionals with application to proximal causal inference', *International Conference on Artificial Intelligence and Statistics* pp. 7210–7239.

Hernán, M. A., Hernández-Díaz, S. & Robins, J. M. (2004), 'A structural approach to selection bias', *Epidemiology* pp. 615–625.

Hernán, M. A. & Robins, J. M. (2020), *Causal inference: what if*, Boca Raton: Chapman & Hall/CRC.

Hitchings, M. D., Ranzani, O. T., Dorion, M., D'Agostini, T. L., de Paula, R. C., de Paula, O. F. P., de Moura Villela, E. F., Torres, M. S. S., de Oliveira, S. B., Schulz, W. et al. (2021), 'Effectiveness of ChAdOx1 vaccine in older adults during SARS-CoV-2 Gamma variant circulation in São Paulo', *Nature communications* **12**(1), 6220.

Hudgens, M. G. & Halloran, M. E. (2006), 'Causal vaccine effects on binary postinfection outcomes', *Journal of the American Statistical Association* **101**(473), 51–64.

Hudgens, M. G. & Halloran, M. E. (2008), 'Toward causal inference with interference', *Journal of the American Statistical Association* **103**(482), 832–842.

Imbens, G., Kallus, N. & Mao, X. (2021), 'Controlling for unmeasured confounding in panel data using minimal bridge functions: From two-way fixed effects to factor models', *arXiv:2108.03849v1 [stat.ME]* .

Israel, A., Merzon, E., Schäffer, A. A., Shenhar, Y., Green, I., Golan-Cohen, A., Ruppin, E., Magen, E. & Vinker, S. (2021), 'Elapsed time since BNT162b2 vaccine and risk of SARS-CoV-2 infection: test negative design study', *BMJ* **375**.

Jackson, L. A., Jackson, M. L., Nelson, J. C., Neuzil, K. M. & Weiss, N. S. (2006), 'Evidence of bias in estimates of influenza vaccine effectiveness in seniors', *International Journal of Epidemiology* **35**(2), 337–344.

Jackson, M. L., Chung, J. R., Jackson, L. A., Phillips, C. H., Benoit, J., Monto, A. S., Martin, E. T., Belongia, E. A., McLean, H. Q., Gaglani, M. et al. (2017), 'Influenza vaccine effectiveness in the united states during the 2015–2016 season', *New England Journal of Medicine* **377**(6), 534–543.

Jackson, M. L. & Nelson, J. C. (2013), 'The test-negative design for estimating influenza vaccine effectiveness', *Vaccine* **31**(17), 2165–2168.

Jewell, N. P., Dufault, S., Cutcher, Z., Simmons, C. P. & Anders, K. L. (2019), 'Analysis of cluster-randomized test-negative designs: cluster-level methods', *Biostatistics* **20**(2), 332–346.

Kallus, N., Mao, X. & Uehara, M. (2022), 'Causal inference under unmeasured confounding with negative controls: A minimax learning approach', *arXiv:2103.14029v4 [stat.ML]* .

Krammer, F. (2019), 'The human antibody response to influenza a virus infection and vaccination', *Nature Reviews Immunology* **19**(6), 383–397.

Leung, J., Harpaz, R., Molinari, N.-A., Jumaan, A. & Zhou, F. (2011), 'Herpes zoster incidence among insured persons in the united states, 1993–2006: evaluation of impact of varicella vaccination', *Clinical Infectious Diseases* **52**(3), 332–340.

Lipsitch, M., Jha, A. & Simonsen, L. (2016), 'Observational studies and the difficult quest for causality: lessons from vaccine effectiveness and impact studies', *International Journal of Epidemiology* **45**(6), 2060–2074.

Lipsitch, M., Tchetgen Tchetgen, E. & Cohen, T. (2010), 'Negative controls: a tool for detecting confounding and bias in observational studies', *Epidemiology* **21**(3), 383.

Miao, W., Geng, Z. & Tchetgen Tchetgen, E. (2018), 'Identifying causal effects with proxy variables of an unmeasured confounder', *Biometrika* **105**(4), 987–993.

Miao, W., Shi, X. & Tchetgen Tchetgen, E. (2020), 'A confounding bridge approach for double negative control inference on causal effects', *arXiv:1808.04945v3 [stat.ME]* .

Moline, H. L., Whitaker, M., Deng, L., Rhodes, J. C., Milucky, J., Pham, H., Patel, K., Anglin, O., Reingold, A., Chai, S. J. et al. (2021), 'Effectiveness of COVID-19 vaccines in preventing hospitalization among adults aged $\geq$ 65 years—COVID-NET, 13 states, February–April 2021', *Morbidity and Mortality Weekly Report* **70**(32), 1088.

Newey, W. K. & Powell, J. L. (2003), 'Instrumental variable estimation of nonparametric models', *Econometrica* **71**(5), 1565–1578.

Olson, S. M., Newhams, M. M., Halasa, N. B., Price, A. M., Boom, J. A., Sahni, L. C., Pannaraj, P. S., Irby, K., Walker, T. C., Schwartz, S. P. et al. (2022), 'Effectiveness of BNT162b2 vaccine against critical Covid-19 in adolescents', *New England Journal of Medicine* .

Patel, M. M., Jackson, M. L. & Ferdinands, J. (2020), 'Postlicensure evaluation of COVID-19 vaccines', *Journal of the American Medical Association* **324**(19), 1939–1940.

Schnitzer, M. E. (2022), 'Estimands and estimation of covid-19 vaccine effectiveness under the test-negative design: connections to causal inference', *Epidemiology* **33**(3), 325.

Shi, X., Miao, W., Hu, M. & Tchetgen Tchetgen, E. (2023), 'Theory for identification and inference with synthetic controls: A proximal causal inference framework', *arXiv:2108.13935v4 [stat.ME]* .

Shi, X., Miao, W. & Tchetgen Tchetgen, E. (2020), 'A selective review of negative control methods in epidemiology', *Current Epidemiology Reports* pp. 1–13.

Sullivan, S. G., Feng, S. & Cowling, B. J. (2014), 'Potential of the test-negative design for measuring influenza vaccine effectiveness: a systematic review', *Expert Review of Vaccines* **13**(12), 1571–1591.

Sullivan, S. G., Tchetgen Tchetgen, E. & Cowling, B. J. (2016), 'Theoretical basis of the test-negative study design for assessment of influenza vaccine effectiveness', *American Journal of Epidemiology* **184**(5), 345–353.

Tchetgen Tchetgen, E. & VanderWeele, T. J. (2012), 'On causal inference in the presence of interference', *Statistical Methods in Medical Research* **21**(1), 55–75.

Tchetgen Tchetgen, E., Ying, A., Cui, Y., Shi, X. & Miao, W. (2020), 'An introduction to proximal causal learning', *arXiv:2009.10982v1 [stat.ME]* .

Thompson, M. G., Stenehjem, E., Grannis, S., Ball, S. W., Naleway, A. L., Ong, T. C., DeSilva, M. B., Natarajan, K., Bozio, C. H., Lewis, N. et al. (2021), 'Effectiveness of COVID-19 vaccines in ambulatory and inpatient care settings', *New England Journal of Medicine* **385**(15), 1355–1371.

Van der Vaart, A. (2000), *Asymptotic statistics*, Vol. 3, Cambridge university press.

Wang, B., Dufault, S. M., Small, D. S. & Jewell, N. P. (2023), 'Randomization inference for cluster-randomized test-negative designs with application to dengue studies: Unbiased estimation, partial compliance, and stepped-wedge design', *Annals of Applied Statistics* (in press).

Ying, A., Miao, W., Shi, X. & Tchetgen Tchetgen, E. (2023), 'Proximal causal inference for complex longitudinal studies', *Journal of the Royal Statistical Society Series B: Statistical Methodology* (in press).