Biometrika (2022), **xx**, x, *p.* 1
Printed in Great Britain

Identified vaccine efficacy for binary post-infection outcomes under misclassification without monotonicity

By R. Trangucci
Department of Statistics, University of Michigan, Ann Arbor MI, USA
tranguc@umich.edu

Y. Chen
Department of Statistics, University of Michigan, Ann Arbor MI, USA
ychenang@umich.edu

J. Zelner
Department of Epidemiology & Center for Social Epidemiology and Population Health, University of Michigan School of Public Health, Ann Arbor, MI, USA
jzelner@umich.edu

Summary

Despite the importance of vaccine efficacy against post-infection outcomes like transmission or severe illness, these estimands are unidentifiable, even under strong assumptions that are rarely satisfied in real-world trials. We develop a novel method to nonparametrically point identify these principal effects while eliminating the monotonicity assumption and allowing for measurement error. Furthermore, our results allow for multiple treatments, and are general enough to be applicable outside of vaccine efficacy. Our method relies on the fact that many vaccine trials are run across geographically disparate sites, and measure biologically-relevant categorical pretreatment covariates. We show that our method can be applied to a variety of clinical trial settings where vaccine efficacy against infection and a post-infection outcome can be jointly inferred. This can yield new insights from existing vaccine efficacy trial data and will aid researchers in designing new multi-arm clinical trials.

1. Introduction

Vaccine efficacy against binary post-infection outcomes like symptoms, severe illness or death is of the utmost importance for public health policy makers; it helps policy makers optimize vaccination programs, communicate with the public, allocate scarce resources, and guide future pharmaceutical therapeutic development (Lipsitch & Kahn, 2021). Causal inference for post-infection outcomes was initially developed for continuous outcomes in Gilbert et al. (2003), which relies on the principal stratification framework introduced by Frangakis & Rubin (2002). The methodology was further developed in Jemiai

et al. (2007); Shepherd et al. (2006, 2007). The estimand for binary outcomes, however, was first developed by Hudgens & Halloran (2006). Unfortunately, vaccine efficacy against binary post-infection outcomes is not identifiable, even under the assumption that vaccine efficacy against infection is non-negative almost-surely (monotonicity). Moreover, the method requires that both infection and post-infection outcomes are perfectly measured, and requires sensitivity analyses to derive plausible bounds. These complexities may have stymied its use. For example, the World Health Organization's *Evaluation of Influenza Vaccine Effectiveness* notes "Measures of outcome severity, such as duration, subsequent hospitalization (particularly for outpatient outcomes), or death, may be useful for assessing whether influenza vaccine reduces severity of outcomes in the vaccinated (*although this is complicated to estimate*)" (World Health Organization (2017), emphasis ours).

We develop novel methodology to point identify vaccine efficacy against binary post-infection outcomes without assuming monotonicity while allowing infection and post-infection outcomes to be misclassified. Our framework immediately generalizes to multiple treatments as we will show. We build on literature for identifying principal stratum effects with covariates (Rubin, 2006; Ding et al., 2011; Jiang et al., 2016), on using covariates to hone large-sample nonparametric bounds (Zhang & Rubin, 2003; Grilli & Mealli, 2008; Long & Hudgens, 2013), and on identifying causal estimands under unmeasured confounding (Miao et al., 2018; Shi et al., 2020). Our method also fits into recent literature on inferring causal estimands under measurement error (Jiang & Ding, 2020) and on identification of latent variable models (Ouyang & Xu, 2022). We show that our method can be used to design randomized trials for comparison of multiple vaccines against a control, which will be a necessity for public health agencies in future pandemics as well as during the COVID-19 pandemic. As noted by several authors, vaccine efficacy against post-infection outcomes is mathematically analogous to the widely-studied survivor average treatment effects (Ding et al., 2011; Tchetgen Tchetgen, 2014; Ding & Lu, 2017), so our methodology can be readily used outside the domain of vaccine efficacy.

2. Principal stratification for vaccine efficacy

Borrowing notation from Hudgens & Halloran (2006), suppose we observe the following triplets for each participant in a vaccination trial: (S_i, Y_i, Z_i), where Z_i is a categorical variable with N_z levels representing observed treatment status for individual i, S_i is observed binary infection status, and Y_i is observed binary post-infection outcome. Further, let the vector $(S(z), Y(z, S(z)))$ be the counterfactual infection status and postinfection outcome that would be observed under vaccination status z. This encodes our first assumption:

Assumption 1 *(SUTVA).* There is only one version of each treatment, and counterfactual outcomes are a function of only a unit's respective treatment status, z.

SUTVA can be satisfied for vaccine efficacy trials by restrictions on participants and recruitment (Gilbert et al., 2003). Note that one cannot have a post-infection outcome if one is not infected. This leads to $Y(z, 0)$ being undefined for all z, and is denoted as $Y(z, 0) = \star$. Y is defined as a binary variable only when $S(z) = 1$, or, equivalently, $Y(z, 1)$. Given Assumption 1, the observed variables can be defined in terms of the counterfactual variables:

$$S_i = S(z)\mathbb{1}(Z_i = z), \quad Y_i = Y(z, S(z))\mathbb{1}(Z_i = z). \tag{1}$$

We can model the joint distribution of the observed data (S_i, Y_i) in terms of the joint distribution over $(S(z), Y(z, S(z)))$. Let $p_{syz} = P(S_i = s, Y_i = y \mid Z_i = z)$ with all p_{01z} undefined. Let the set of principal strata, $\mathcal{S}$, be defined as the set of all possible ordered counterfactual infection outcomes, or

$$\mathcal{S} = \{(S(z_1), S(z_2), \ldots, S(z_{N_z})) \mid S(z_j) \in \{0,1\}\},$$

and let the principal stratum S^{P_0} be an element of $\mathcal{S}$. Let $\theta_u = P(S^{P_0} = u)$ where $u \in \{0,1\}^{N_z}$ and let $\phi_e^u = P(Y(z_1) = e_1, \ldots, Y(z_{N_z}) = e_{N_z} | S^{P_0} = u)$ such that $e_j = \star$ where $u_j = 0$ and $e_j \in \{0,1\}$ only for elements j for which $u_j = 1$. Let $\beta_j^u = \sum_{e|e_j=1} \phi_e^u \equiv P(Y(z_j) = 1 \mid S^{P_0} = u)$, where β_j^u is only defined for j such that $u_j = 1$.

Our second assumption is that treatment assignment is unconfounded with potential outcomes:

Assumption 2 *(Unconfounded treatment assignment).* Treatment assignment is independent of principal stratum, or $S^{P_0} \perp\!\!\!\perp Z$ and, conditional on principal stratum, treatment assignment is independent of counterfactual post-infection outcomes, or $(Y(z_1), \ldots, Y(z_{N_z})) \perp\!\!\!\perp Z \mid S^{P_0}$.

Then our definition for the vaccine efficacy estimands are:

Definition 1 (Vaccine efficacy against infection z_j versus z_k).

$$\mathrm{VE}_{S,jk} = 1 - P(S(z_j))/P(S(z_k)),$$

and

Definition 2 (Vaccine efficacy against post-infection outcome Y).

$$\mathrm{VE}_{I,jk}^u = 1 - P(Y(z_j) | S^{P_0} = u)/P(Y(z_k) | S^{P_0} = u).$$

$\mathrm{VE}_{I,jk}^u$ is a principal effect as defined in Frangakis & Rubin (2002) because it is conditional on a principal stratum u. This causal effect is only defined when u_j and $u_k = 1$. For example, when $N_z = 3$, there are 8 principal strata, three of which would admit comparisons between two treatments: $(1,1,0), (0,1,1), (1,0,1)$, and one of which would allow for comparisons between all three treatments: $(1,1,1)$. We call the stratum $S^{P_0} = \{1\}^{N_z}$ the "always-infected" stratum.

With the expanded set of principal strata, we could define alternative measures of vaccine efficacy. Suppose we were interested in the reduction in the risk of post-infection outcome under z_3 versus z_2, relative to the risk of infection under treatment z_1. Then we could define the following causal estimand:

$$\frac{P(Y(z_2) = 1 \mid S^{P_0} = (1,1,1)) - P(Y(z_3) = 1 \mid S^{P_0} = (1,1,1))}{P(Y(z_1) = 1 \mid S^{P_0} = (1,1,1))} = \mathrm{VE}_{I,31}^{(1,1,1)} - \mathrm{VE}_{I,21}^{(1,1,1)}, \quad (2)$$

which captures this effect. For instance, this comparison might be of interest in a placebo-controlled randomized trial with two competing vaccines; in Monto et al. (2009) a vaccine containing live-attenuated virus is compared to that with an inactivated virus and a placebo group.

2.1. Basic N_z-treatment model

The most basic model with N_z greater than or equal to 2 treatments is an instructive example to solidify the concepts and challenges in inferring post-infection outcome vaccine efficacy estimands. The complete set of observed data comprises triplets $(S_i, Y_i \mid Z_i)$, for each individual i, $i \in \{1, \ldots, n\}$, while the set of treatments is $\{z_1, \ldots, z_{N_z}\}$. By As-

sumption 1, $Z_i \in \{z_1, \ldots, z_{N_z}\}$. The observational model for these data implied by Assumptions 1 to 2 is a stratified multinomial model:

$$(n_{0\star}(z), n_{10}(z), n_{11}(z)) \sim \text{Multinomial}(n(z) \mid p_{0*z}, p_{10z}, p_{11z}),\ z \in \{z_1, \ldots, z_{N_z}\}, \quad (3)$$

where $n_{sy}(z)$ and $n(z)$ are defined as:

$$n_{sy}(z) = \sum_{i=1}^{n} \mathbb{1}(S_i = s)\, \mathbb{1}(Y_i = y)\, \mathbb{1}(Z_i = z), \quad n(z) = \sum_{i=1}^{n} \mathbb{1}(Z_i = z).$$

We define the observed probabilities as

$$p_{1yj} = \textstyle\sum_{u|u\in\mathcal{S},u_j=1} \theta_u (\beta_j^u)^y (1-\beta_j^u)^{1-y}, \quad p_{0*j} = 1 - p_{10z} - p_{11z}, \quad (4)$$

where $u \in \{0,1\}^{N_z}$. The model parameters are not identifiable, in the sense of Rothenberg (1971) (defined in the Supplementary Materials). That the parameters are not identifiable is a consequence of having $2^{N_z} - 1 + \sum_{j=1}^{N_z} \binom{N_z}{j}(2^j - 1)$ model parameters, but only $2N_z$ degrees of freedom in the observed data distribution.

The canonical two-arm vaccine efficacy trial is an instructive example of the structure of the identifiable model parameter subspace. Let $N_z = 2$, and to accord with the notation in Hudgens & Halloran (2006), let the group z_1 be the vaccinated group, or $z_1 = 1$, and let group z_2 be the placebo group, or $z_2 = 0$. Let $u \in \{(0,0),(1,0),(0,1),(1,1)\}$. Note that under monotonicity, $P(S^{P_0} = (1,0)) = 0$. We use ideas from Gustafson (2015) to define the identifiable model parameter subspace. The map from the model parameters to the observable probabilities is:

$$\begin{aligned}
p_{110} &= \theta_{(0,1)}\beta_2^{(0,1)} + \theta_{(1,1)}\beta_2^{(1,1)}, & p_{111} &= \theta_{(1,0)}\beta_1^{(1,0)} + \theta_{(1,1)}\beta_1^{(1,1)} \\
p_{100} &= \theta_{(0,1)}\left(1-\beta_2^{(0,1)}\right) + \theta_{(1,1)}\left(1-\beta_2^{(1,1)}\right), & p_{101} &= \theta_{(1,0)}\left(1-\beta_1^{(1,0)}\right) + \theta_{(1,1)}\left(1-\beta_1^{(1,1)}\right),
\end{aligned}$$

where $\beta_1^{(1,1)} = \phi_{11}^{(1,1)} + \phi_{10}^{(1,1)}, \beta_2^{(1,1)} = \phi_{11}^{(1,1)} + \phi_{01}^{(1,1)}$. Plainly, the joint distribution of the observed data has only 4 independent quantities, but the probability model has 8 parameters. Thus, the model parameters are unidentified. We can define the identifiable subspace in terms of the observable probabilities and the unidentified parameters $\phi_{10}^{(1,1)}, \beta_2^{(0,1)}, \beta_1^{(1,0)}, \theta_{(1,1)}$. This reparameterization is not unique. In fact, the structure of the model is such that no single model parameter is identifiable without further restrictions like monotonicity or auxiliary information, like that investigated in Jiang et al. (2016) and Ding et al. (2011). The map to the identifiable subspace is

$$\begin{aligned}
\theta_{(0,1)} &= p_{1+0} - \theta_{(1,1)}, \quad \phi_{11}^{(1,1)} = \frac{p_{111} - \beta_1^{(1,0)} p_{1+1}}{\theta_{(1,1)}} - \phi_{10}^{(1,1)} + \beta_1^{(1,0)} \\
\theta_{(1,0)} &= p_{1+1} - \theta_{(1,1)}, \quad \phi_{01}^{(1,1)} = \frac{p_{110} + \beta_1^{(1,0)} p_{1+1} - \beta_2^{(0,1)} p_{1+0} - p_{111}}{\theta_{(1,1)}} + \phi_{10}^{(1,1)} + \beta_2^{(0,1)} - \beta_1^{(1,0)}.
\end{aligned}$$

The causal estimand, $\text{VE}_{I,12}^{(1,1)}$, does not depend on $\phi_{10}^{(1,1)}$ and depends on only the unidentified parameters $\beta_2^{(0,1)}, \beta_1^{(1,0)}, \theta_{(1,1)}$

$$1 - \frac{\phi_{11}^{(1,1)} + \phi_{10}^{(1,1)}}{\phi_{11}^{(1,1)} + \phi_{01}^{(1,1)}} = 1 - \frac{p_{111} + \beta_1^{(1,0)}(\theta_{(1,1)} - p_{1+1})}{p_{110} + \beta_2^{(0,1)}(\theta_{(1,1)} - p_{1+0})}. \quad (5)$$

Note that if one assumes monotonicity, $\theta_{(1,1)}$ is identifiable, $\beta_1^{(1,0)} = 0$ and the causal estimand depends on only one unidentifiable parameter, $\beta_2^{(0,1)}$, or the probability of a post-infection outcome in the control group for the principal stratum where $(S(1) = 0, S(0) = 1)$. Even if monotonicity is a reasonable assumption, it is not an assumption that is made when assessing vaccine efficacy against infection. Accordingly, two separate analyses must be performed to assess the two efficacy estimands. The price for two-step procedures is two-fold: to the extent there is shared information they are less statistically efficient, and, because two sets of assumptions are needed, it may be harder to communicate results to stakeholders.

While lack of identifiability is not in and of itself inherently objectionable, it is worthwhile to examine conditions under which identifiability of the causal estimand is achieved. If these conditions are applicable in real-world trials, we can design such trials so as to jointly learn vaccine efficacy against infection and vaccine efficacy against post-infection outcomes.

In the subsequent section we show that vaccine efficacy against post-infection outcomes is identifiable under reasonable assumptions, and that these results hold for more than two treatments, which is important in scenarios where several vaccines are under study, as is the case for the study in Monto et al. (2009), and in COVID-19 where there is interest in comparing the efficacy of several approved vaccines against severe illness or death (Tenforde et al., 2021).

2.2. *Identifiability under multiple study sites and a categorical covariate*

As shown in Section 2.1, model (4) is not identifiable because the number of parameters is greater than the number of degrees of freedom in the observed probabilities. We can identify our model by expanding the number of measured covariates while limiting the growth in the number of parameters. We do so by employing reasonable conditional independence assumptions. Our solution is motivated by the structure of vaccine efficacy trials: many vaccine efficacy trials are run at multiple geographically disparate sites, and typically include pretreatment biological or demographic attributes that may be related to the principal strata and the post-infection outcome.

The model identifiability result is novel in causal inference because we do not require any monotonicity assumptions, it holds for an arbitrary number of treatments, and we do not make distributional assumptions on the secondary outcome $Y(z_j)$. This means that Theorem 1 is applicable to the identifiability of any principal stratification model where the intermediate outcome is binary. Moreover, the structure of the proof lends itself to further generalization for categorical or multivariate binary outcomes, making the method applicable to noncompliance in multi-arm studies without requiring an exclusion restriction.

If the study is run across multiple sites, like different healthcare systems, each participant is associated with a study site, indicated by a categorical variable R_i. Then if the study also measures a principal-stratum-relevant pretreatment categorical covariate A_i for each participant we can use the joint variation in covariate values across study sites to identify the principal strata proportions by study site. An example of a relevant pretreatment covariate in a vaccine efficacy study is pre-vaccination, pre-season antibody concentrations, or a measurement of the health of the immune system. Then we may use the variation in principal strata proportions between study sites to identify the distribution of post-infection potential outcomes.

Let R_i take values from 1 to N_r and A_i take values from 1 to N_a. Further, suppose that the probability model satisfies the following two assumptions:

Assumption 3 *(Covariate conditional independence by site).* A is conditionally independent of the study site and treatment receipt given the principal stratum, or $A \perp\!\!\!\perp R, Z \mid S^{P_0}$.

Assumption 4 *(Causal Homogeneity).* Conditional on principal stratum S^{P_0} and A, the potential outcomes $(Y(z_1), \ldots, Y(z_{N_z}))$ are independent of R, or $(Y(z_1), \ldots, Y(z_{N_z})) \perp\!\!\!\perp R \mid S^{P_0}, A$.

Let $\theta_u^r = P(S^{P_0} = u \mid R_i = r)$, $a_k^u = P(A = k \mid S^{P_0} = u)$, and $\beta_{j,k}^u = P(Y(z_j) = 1 \mid S^{P_0} = u, A = k)$. Let $p_{sykzr} = P(S_i = s, Y_i = y, A_i = k \mid Z_i = z, R_i = r)$ be the observable probabilities Under assumptions 1 to 4, the probability model can be defined:

$$\begin{aligned}&(n_{0\star 1}(z,r), n_{101}(z,r), n_{111}(z,r), \ldots, n_{0\star N_a}(z,r), n_{10N_a}(z,r), n_{11N_a}(z,r)) \sim \\ &\text{Multinomial}(n(z,r) \mid p_{0\star 1zr}, p_{101zr}, p_{111zr}, \ldots, p_{0\star N_a zr}, p_{10N_a zr}, p_{11N_a zr}), \\ &z \in \{1, \ldots, N_z\},\ r \in \{1, \ldots, N_r\}.\end{aligned} \tag{6}$$

Let $n_{syk}(z,r)$ be defined as:

$$n_{syk}(z,r) = \sum_{i=1}^{n} \mathbb{1}(S_i = s)\, \mathbb{1}(Y_i = y)\, \mathbb{1}(Z_i = z)\, \mathbb{1}(R_i = r)\, \mathbb{1}(A_i = k),$$

and

$$p_{1ykjr} = \textstyle\sum_{u \mid u \in \mathcal{S}, u_j = 1} a_k^u \theta_u^r (\beta_{j,k}^u)^y (1 - \beta_{j,k}^u)^{1-y}, \quad p_{0\star kjr} = \textstyle\sum_{u \mid u \in \mathcal{S}, u_j = 0} a_k^u \theta_u^r. \tag{7}$$

The number of parameters in the model for infection and covariates is $N_r(2^{N_z} - 1) + 2^{N_z}(N_a - 1)$ while the number of parameters in the post-infection outcome model is $N_a(2^{N_z} - 1)$. This definition omits the parameters ϕ_e^u which are fundamentally unidentifiable given that they correspond to unobservable outcomes. The number of degrees of freedom in the observational model is $N_r N_z(2N_a - 1)$. This shows that for a fixed N_z, N_r and N_a can be large enough so that the observed probabilities are more numerous than the parameters. Thus, we may be able to identify the model parameters.

In order to develop a framework for multiple treatment settings, we may define an ordering among the principal strata. Given that the principal strata are binary strings of length N_z, a natural ordering among the strata are the base-10 representations of the strings. We define an operator and its inverse to map from the integers to principal stratum and vice-versa:

Definition 3 (Base-10 to binary map). *Let the operator ϖ_m be defined as $\varpi_m(\cdot) : j \to \{0,1\}^m, j \in \mathbb{N}, j \le 2^m - 1$ with elements $\varpi_m(j)_i \in \{0,1\}$, so $\varpi_m(j)$ is the base-2 representation of j with m digits represented as a binary m-vector. Let the inverse operator $\varpi_m^{-1}(\cdot) : \{0,1\}^m \to j$, or the binary to base-10 conversion.*

For example, $\varpi_3(4) = (0,0,1)$, and $\varpi_5(4) = (0,0,1,0,0)$, and $\varpi_5^{-1}((0,0,1,0,0)) = 4$.

Now we can define two matrices which will be common to all models no matter the number of treatments. Let the matrix $P_{N_z}(A \mid S^{P_0})$ in $\mathbb{R}^{N_a \times 2^{N_z}}$ encode the distributions $A \mid S^{P_0}$ with $(i,j)^{\text{th}}$ element $P_{N_z}(A = i \mid S^{P_0} = \varpi_{N_z}(j-1))$. Let the matrix $P_{N_z}(S^{P_0} \mid R)$ in $\mathbb{R}^{2^{N_z} \times N_r}$ encode the distribution $S^{P_0} \mid R$ with $(i,j)^{\text{th}}$ element $P_{N_z}(S^{P_0} = \varpi_{N_z}(i-1) \mid R = j)$. Before introducing Theorem 1, let us introduce the concept of the Kruskal rank of a matrix.

Definition 4 (Kruskal rank). *Let the Kruskal rank of a matrix $B \in \mathbb{R}^{I \times R}$ be $k_B \in [0, 1, 2, \dots)$, and let k_B be the maximum integer such that every set of k_B columns of B are linearly independent.*

Kruskal rank is stricter than matrix rank. To see why, consider a matrix with R columns of which two are repeated. At most the rank can be $R-1$, but Kruskal rank can be at most 1. Kruskal rank equals column rank if a matrix is full column rank.

We present the first of two primary results of our paper.

Theorem 1. *Let the number of treatments be $N_z \geq 2$, and the number of principal strata be 2^{N_z}. Suppose that Assumptions 1 to 4 hold. Then if the matrix $P_{N_z}(A \mid S^{P_0})$ is Kruskal rank greater than or equal to $2^{N_z} - 1$, and the rank of matrix $P_{N_z}(S^{P_0} \mid R)$ is 2^{N_z} then the univariate counterfactual distributions $P(Y(z_j) = 1 \mid S^{P_0} = u, A = k)$ are identifiable, as are $P(S^{P_0} = u \mid R = r)$, and $P(A = k \mid S^{P_0} = u)$.*

The identifiability of the vaccine efficacy estimands is a by-product of the identifiability of the conditional counterfactual distributions for $Y(z_j) \mid S^{P_0}, A$. Furthermore the identifiability of the conditional counterfactual distributions $P(Y(z_j) = 1 \mid S^{P_0} = u, A = k)$ allows for causal effect heterogeneity by covariate A.

Definition 5 (Conditional VE against post-infection outcome Y).

$$\mathrm{VE}^{u}_{I,jk}(k) = 1 - \frac{P(Y(z_j) | S^{P_0} = u, A = k)}{P(Y(z_k) | S^{P_0} = u, A = k)}$$

Continuing the $N_z = 3$ example, where there are 8 principal strata, there are 6 conditional post-infection outcome vaccine efficacy estimands for each value of A.

As these estimands are conditional over A, we can infer the estimands in Equation (2) that marginalize over the population distribution of A. This population distribution is identifiable by Theorem 1.

Corollary 1. *By the conditions set forth in Theorem 1, $P(A = k \mid S^{P_0} = u)$ and $P(Y(z_j) = 1 \mid S^{P_0} = u, A = k)$ are identifiable for $u \in \mathcal{S}$ and $k \in \{1, \dots, N_a\}$. Let $P(Y(z_j) = 1 \mid S^{P_0} = u) = \sum_k P(A = k \mid S^{P_0} = u) P(Y(z_j) = 1 \mid S^{P_0} = u, A = k)$. Then for $u \in \mathcal{S}$, $P(Y(z_j) = 1 \mid S^{P_0} = u), P(A = k \mid S^{P_0} = u)$, and $P(Y(z_j) = 1 \mid S^{P_0} = u, A = k)$ are identifiable.*

Our method is related to methods in Jiang et al. (2016) and Ding et al. (2011). Ding et al. (2011) addresses problems of identifiability in survivor average treatment effects, which is mathematically analogous to vaccine efficacy for post-infection outcomes, by measuring covariates that are related to the principal strata. Jiang et al. (2016) identifies principal causal effects in binary surrogate endpoint evaluations. Despite not being mathematically identical to vaccine efficacy, binary surrogacy endpoint evaluation is ultimately a problem in identification of principal causal effects. The proof of Theorem 1 is shown in the Supplementary Materials, but it relies on two proof techniques. The first is developed in Miao et al. (2018) and extended in Shi et al. (2020), and the second is an extension of the Three-Way Array Decomposition Uniqueness Theorem (Kruskal, 1977). Our extension is similar to that in Allman et al. (2009), but our results allow for strict identifiability, whereas Allman et al. (2009) only shows identifiability up to a permutation. Strict identifiability is necessary for our application to principal stratification. Most importantly, the proof does not encode any restrictions on the distribution of secondary outcomes, otherwise known in our case as the post-infection outcomes. This makes the result applicable to categorical or continuous post-infection outcomes, and, more broadly, to principal stratification problems outside the scope of vaccine efficacy.

2.3. *Models and sensitivity analyses*

The post-infection outcome models can be formulated as logistic regressions:

$$\log \frac{P(Y(z_j)=1 \mid S^{P_0}=u, A=k)}{P(Y(z_j)=0 \mid S^{P_0}=u, A=k)} = \alpha_j^u + \delta_{j,k}^u,\ \beta_{j,k}^u = \frac{e^{\alpha_j^u+\delta_{j,k}^u}}{1+e^{\alpha_j^u+\delta_{j,k}^u}},\quad \delta_{j,1}^u = 0\ \forall\ j,\ u.$$

Deviations from Assumption 4 can be encoded as an additive term γ_r^u capturing heterogeneity between study sites:

$$\log \frac{P(Y(z_j)=1 \mid S^{P_0}=u, A=k, R=r)}{P(Y(z_j)=0 \mid S^{P_0}=u, A=k, R=r)} = \alpha_j^u + \delta_{j,k}^u + \varepsilon_r^u,\quad \varepsilon_r^u \sim \text{Normal}(0, (\tau_\varepsilon^u)^2).$$

We can fix τ_γ^u to several values for sensitivity analysis, as developed in Jiang et al. (2016).

We may write the probability model in Equation (7) as two multinomial regressions, given Assumption 3 that $A \perp\!\!\!\perp R, Z \mid S^{P_0}$. It is straightforward to include pretreatment covariates $X \in \mathbb{R}^M$ into both regression models:

$$\log \frac{P(S^{P_0}=u \mid R=r, X=x)}{P(S^{P_0}=u_0 \mid R=r, X=x)} = \mu_u^r + x^T\eta_u,\ \log \frac{P(A=k \mid S^{P_0}=u, X=x)}{P(A=k_0 \mid S^{P_0}=u, X=x)} = \nu_k^u + x^T\gamma_k,$$

where

$$\theta_u^{r,x} = \frac{e^{\mu_u^r + x^T\eta}}{\sum_{w\in\mathcal{S}} e^{\mu_w^r + x^T\eta}},\ \mu_{u_0}^r = 0\ \forall\ r, \qquad a_k^{u,x} = \frac{e^{\nu_k^u + x^T\gamma}}{\sum_{m=1}^{N_a} e^{\nu_m^u + x^T\gamma}},\ \nu_{k_0}^u = 0\ \forall\ u.$$

Note η_u and γ_k are in $\mathbb{R}^M$ with η_{u_0}, γ_{k_0} each as the M-dimensional zero vector. This leads to a tidy representation of the log-odds of belonging to stratum u vs. u_0 conditional on $A=k, R=r, X=x$:

$$\log \frac{P(S^{P_0}=u \mid A=k, R=r, X=x)}{P(S^{P_0}=u_0 \mid A=k, R=r, X=x)} = \mu_u^r + \nu_k^u - \nu_k^{u_0} + x^T\eta_u.$$

If we suspect deviations from Assumption 3, we can add an interaction between A and R:

$$\log \frac{P(A=k \mid R=r, S^{P_0}=u)}{P(A=k_0 \mid R=r, S^{P_0}=u)} = \nu_k^u + \epsilon_{k,r}^u,\quad \epsilon_{k,r}^u \sim \text{Normal}(0, (\tau_\epsilon^u)^2).$$

2.4. *Randomized trial with two treatments*

Our result subsumes the canonical two-treatment randomized trial setting, i.e. $N_z = 2$, which we state as a corollary. In this case, the probability model is completely defined as:

$$p_{11k0r} = a_k^{(0,1)}\theta_{(0,1)}^r\beta_{2,k}^{(0,1)} + a_k^{(1,1)}\theta_{(1,1)}^r\beta_{2,k}^{(1,1)}, \qquad p_{11k1r} = a_k^{(1,0)}\theta_{(1,0)}^r\beta_{1,k}^{(1,0)} + a_k^{(1,1)}\theta_{(1,1)}^r\beta_{1,k}^{(1,1)}$$

$$p_{10k0r} = a_k^{(0,1)}\theta_{(0,1)}^r(1-\beta_{2,k}^{(0,1)}) + a_k^{(1,1)}\theta_{(1,1)}^r(1-\beta_{2,k}^{(1,1)}), \quad p_{10k1r} = a_k^{(1,0)}\theta_{(1,0)}^r(1-\beta_{1,k}^{(1,0)}) + a_k^{(1,1)}\theta_{(1,1)}^r(1-\beta_{1,k}^{(1,1)}).$$

Then the matrices $P_2(A \mid S^{P_0}), P_2(S^{P_0} \mid R)$ are defined

$$P_2(A \mid S^{P_0}) = \begin{bmatrix} a_1^{(0,0)} & a_1^{(1,0)} & a_1^{(0,1)} & a_1^{(1,1)} \\ a_2^{(0,0)} & a_2^{(1,0)} & a_2^{(0,1)} & a_2^{(1,1)} \\ \vdots & \vdots & \vdots & \vdots \\ a_{N_a}^{(0,0)} & a_{N_a}^{(1,0)} & a_{N_a}^{(0,1)} & a_{N_a}^{(1,1)} \end{bmatrix},\ P_2(S^{P_0} \mid R) = \begin{bmatrix} \theta_{(0,0)}^{r_1} & \theta_{(0,0)}^{r_2} & \cdots & \theta_{(0,0)}^{r_{N_r}} \\ \theta_{(1,0)}^{r_1} & \theta_{(1,0)}^{r_2} & \cdots & \theta_{(1,0)}^{r_{N_r}} \\ \theta_{(0,1)}^{r_1} & \theta_{(0,1)}^{r_2} & \cdots & \theta_{(0,1)}^{r_{N_r}} \\ \theta_{(1,1)}^{r_1} & \theta_{(1,1)}^{r_2} & \cdots & \theta_{(1,1)}^{r_{N_r}} \end{bmatrix}.$$

Corollary 2 (Identifiability). *Suppose that Assumptions 1 to 4 hold. Then if $P_2(A \mid S^{P_0})$ is Kruskal rank 3 or greater and $P_2(S^{P_0} \mid R)$ is rank 4, the the counterfactual distributions $P(Y(z_j)=i \mid S^{P_0}=(m,n), A=k)$ are identifiable for i,j,m,n each*

in $\{0,1\}$ *and* $A \in \{1,\ldots,N_a\}$, *as are the distributions* $P(S^{P_0} = (m,n) \mid R = r), P(A = k \mid S^{P_0} = (m,n))$, $r \in \{1,\ldots,N_r\}$, $(m,n) \in \{(0,0),(1,0),(1,0),(1,1)\}$.

2.5. Relaxed assumptions and nonparametric bounds on VE_I

Suppose that a trial is run at multiple sites, but cannot measure a covariate A satisfying Assumption 3 for each participant. We can still use variation in principal strata across sites to identify several counterfactual distributions, but we won't be able to identify all model parameters. Instead, we can derive sharp bounds on the post-infection outcome vaccine efficacy and on the proportion of the population in principal stratum $(S(1) = 1, S(0) = 1)$ by study site.

Similar to section 2.1 let $p_{syzr} = P(S_i = s, Y_i = y \mid Z_i = z, R_i = r)$ with all p_{01zr} undefined. Our simplified observational model is now:

$$\begin{aligned}(n_{0\star}(z,r), n_{10}(z,r), n_{11}(z,r)) \sim \mathrm{Multinomial}(n(z,r) \mid p_{0\star zr}, p_{10zr}, p_{11zr}),\\ z \in \{0,1\}, r \in \{1,\ldots,N_r\}.\end{aligned} \tag{8}$$

Let $n_{sy}(z,r)$ and $n(z,r)$ be defined as:

$$n_{sy}(z,r) = \sum_{i=1}^{n} \mathbb{1}(S_i = s)\,\mathbb{1}(Y_i = y)\,\mathbb{1}(Z_i = z)\,\mathbb{1}(R_i = r),\ n(z,r) = \sum_{i=1}^{n} \mathbb{1}(Z_i = z)\,\mathbb{1}(R_i = r).$$

Given that we no longer have A in our probability model, we need to restate Assumption 4:

Assumption 5 *(Causal Homogeneity).* Conditional on principal stratum S^{P_0}, the potential outcomes $Y(1), Y(0)$ is independent of R, or $(Y(1), Y(0)) \perp\!\!\!\perp R \mid S^{P_0}$.

Under Assumption 4, the probability model may be written:

$$\begin{aligned} p_{110r} &= \theta^r_{(0,1)}\beta_2^{(0,1)} + \theta^r_{(1,1)}\beta_2^{(1,1)}, & p_{111r} &= \theta^r_{(1,0)}\beta_1^{(1,0)} + \theta^r_{(1,1)}\beta_1^{(1,1)}\\ p_{100r} &= \theta^r_{(0,1)}(1-\beta_2^{(0,1)}) + \theta^r_{(1,1)}(1-\beta_2^{(1,1)}), & p_{101r} &= \theta^r_{(1,0)}(1-\beta_1^{(1,0)}) + \theta^r_{(1,1)}(1-\beta_1^{(1,1)}).\end{aligned}$$

Theorem 2 (VE without monotonicity). *Suppose that Assumptions 1 to 2 hold as well as Assumption 5. Further, suppose there exist at least* 3 *sites* r_1, r_2, r_3 *such that*

$$\begin{aligned} &p_{1+1r_1}(p_{110r_2}p_{100r_3} - p_{110r_3}p_{100r_2}) + p_{1+1r_2}(p_{100r_1}p_{110r_3} - p_{100r_3}p_{110r_1})\\ &\quad + p_{1+1r_3}(p_{100r_2}p_{110r_1} - p_{100r_1}p_{110r_2}) \neq 0, \end{aligned} \tag{9}$$

$$\begin{aligned} &p_{111r_1}(p_{1+0r_3}p_{1+1r_2} - p_{1+0r_2}p_{1+1r_3}) + p_{111r_2}(p_{1+0r_1}p_{1+1r_3} - p_{1+0r_3}p_{1+1r_1})\\ &\quad + p_{111r_3}(p_{1+0r_2}p_{1+1r_1} - p_{1+0r_1}p_{1+1r_2}) \neq 0. \end{aligned} \tag{10}$$

Then $\beta_2^{(0,1)}, \beta_1^{(1,0)}$ *are identified, and the sharp bounds for* VE_I *are*

$$\mathrm{VE}_I \in \bigcap_{r=1}^{N_r} ({}_l\mathrm{VE}_{Ir}, {}_u\mathrm{VE}_{Ir}),$$

$$({}_l\mathrm{VE}_{Ir}, {}_u\mathrm{VE}_{Ir}) = \begin{cases} \left(1 - \frac{p_{111}+\beta_1^{(1,0)}\left({}_l\theta^r_{(1,1)} - p_{1+1}\right)}{p_{110}+\beta_2^{(0,1)}\left({}_l\theta^r_{(1,1)} - p_{1+0}\right)}, 1 - \frac{p_{111}+\beta_1^{(1,0)}\left({}_u\theta^r_{(1,1)} - p_{1+1}\right)}{p_{110}+\beta_2^{(0,1)}\left({}_u\theta^r_{(1,1)} - p_{1+0}\right)}\right) & \frac{\partial \mathrm{VE}_{Ir}}{\partial\theta} \geq 0\\ \left(1 - \frac{p_{111}+\beta_1^{(1,0)}\left({}_u\theta^r_{(1,1)} - p_{1+1}\right)}{p_{110}+\beta_2^{(0,1)}\left({}_u\theta^r_{(1,1)} - p_{1+0}\right)}, 1 - \frac{p_{111}+\beta_1^{(1,0)}\left({}_l\theta^r_{(1,1)} - p_{1+1}\right)}{p_{110}+\beta_2^{(0,1)}\left({}_l\theta^r_{(1,1)} - p_{1+0}\right)}\right) & \frac{\partial \mathrm{VE}_{Ir}}{\partial\theta} < 0 \end{cases},$$

where

$$\frac{\partial \mathrm{VE}_{Ir}}{\partial \theta} = \beta_1^{(1,0)}\beta_2^{(0,1)}(p_{1+0} - p_{1+1}) - \beta_1^{(1,0)}p_{110} + \beta_2^{(0,1)}p_{111},$$

$$ {}_l\theta_{11}^r = \max\left(p_{1+0r} + p_{1+1r} - 1, \frac{p_{110r} - \beta_2^{(0,1)}p_{1+0r}}{1 - \beta_2^{(0,1)}}, \frac{p_{110r} - \beta_2^{(0,1)}p_{1+0r}}{-\beta_2^{(0,1)}}, \frac{p_{111} - \beta_1^{(1,0)}p_{1+1}}{-\beta_1^{(1,0)}}, 0\right),$$

$$ {}_u\theta_{11}^r = \min(p_{1+0r}, p_{1+1r}).$$

This result is akin to the bounds provided for a 3-arm randomized study in Cheng & Small (2006), though our bounds are limited to the two-arm setting.

These nonparametric bounds are widely applicable to vaccine efficacy studies because there may not exist a covariate that satisfies Assumption 3, but many vaccine efficacy studies are run across multiple geographically disparate sites. Another strength of our methodology is that we derive an inferential benefit from accumulating information between sites because the bounds on VE_I result from an intersection across sites. A downside of these bounds is that they are valid only asymptotically. If we use the plug-in estimators for the bounds, which requires estimators for p_{syz} and $\beta_2^{01}, \beta_1^{10}$, sampling variation can lead to the lower bound being larger than the upper bound for θ_{11r}. If this occurs, we can use the looser lower bound ${}_l\theta_{11}^r = \max(0, p_{1+0} + p_{1+1} - 1)$ which will always respect the upper bound for any estimator of p_{syz}. Another potential issue is that the intersection of the bounds over study sites could be empty. A solution in this scenario is to use the union of the bounds rather than the intersection.

3. Misclassified infections, post-infection outcomes, and covariates

Now suppose that we cannot observe infection or the post-infection outcomes directly, which accords with the reality of most, if not all, vaccination trials. In these trials, infection is usually determined via a viral culture, polymerase-chain-reaction (PCR) test, or viral titer. All of these methods measure infection with error, with varying levels of sensitivity and specificity. For example, PCRs for COVID-19 have very high specificity, but tend to have sensitivities in the range of 0.6 to 0.8 due to variation among patients in how the virus populates the nasal cavity, variation in swab quality, and viral RNA dynamics (Kissler et al., 2021; Wang et al., 2020). Depending on the severity of the post-infection outcome, these outcomes may also be mismeasured. For instance, a high proportion of participants report symptoms in vaccine efficacy studies, despite many of these participants testing negative for the target disease. In the presence of high-sensitivity tests, this necessarily means that specificity of symptoms following infection is below 1.

In order to make our methods applicable to real-world trials, in this section we will extend the ideas in Section 2.2 to scenarios in which infection and post-infection outcomes are measured with error. We will show that both conditional and marginal post-infection outcome vaccine efficacy can be identified, along with the distribution of principal strata and covariate distributions.

Let $\tilde{S}$ be an imperfect measurement of infection status S, and let $\tilde{Y}$ be an imperfect measurement of post-infection outcome Y. Furthermore, let $\mathrm{sn}_S = P(\tilde{S} = 1 \mid S = 1), \mathrm{sp}_S = P(\tilde{S} = 0 \mid S = 0)$ and $\mathrm{sn}_Y = P(\tilde{Y} = 1 \mid Y = 1), \mathrm{sp}_Y = P(\tilde{Y} = 0 \mid Y = 0)$, or the respective sensitivities and specificities for infection and the post-infection outcome. Let the observable probabilities be defined as $q_{syazr} = P(\tilde{S}_i = s, \tilde{Y}_i = y, A_i = a \mid Z_i = z, R_i = r)$. Contrary to the

noiseless observation model in eq. (6), the conditional probability of observing a negative infection test result and a post-infection outcome in the set $\{0,1\}$, or q_{0yazr} for $y \in \{0,1\}$, is well-defined and enters into the observation model below:

$$\begin{aligned}&(\tilde{n}_{001}(z,r), \tilde{n}_{011}(z,r), \tilde{n}_{101}(z,r), \tilde{n}_{111}(z,r), \ldots, \tilde{n}_{00N_a}(z,r), \tilde{n}_{01N_a}(z,r), \tilde{n}_{10N_a}(z,r), \tilde{n}_{11N_a}(z,r)) \sim \\ &\text{Multinomial}(n(z,r) \mid q_{001zr}, q_{011zr}, q_{101zr}, q_{111zr}, \ldots, q_{00N_azr}, q_{01N_azr}, q_{10N_azr}, q_{11N_azr}), \\ &z \in \{1,\ldots,N_z\},\ r \in \{1,\ldots,N_r\}\end{aligned} \tag{11}$$

Let $\tilde{n}_{syk}(z,r)$ be $\sum_{i=1}^{n} \mathbb{1}\left(\tilde{S}_i = s\right) \mathbb{1}\left(\tilde{Y}_i = y\right) \mathbb{1}\left(Z_i = z\right) \mathbb{1}\left(R_i = r\right) \mathbb{1}\left(A_i = k\right)$. The following non-differential measurement error assumptions are common in measurement error models:

Assumption 6 *(Non-differential infection misclassification).* Misclassification is conditionally independent of treatment, principal stratum, and study site or $\tilde{S} \perp\!\!\!\perp Z, S^{P_0}, R, A \mid S$.

Assumption 7 *(Nondifferential outcome misclassification).* Misclassification is conditionally independent of treatment, principal stratum, and study site or $\tilde{Y} \perp\!\!\!\perp Z, S^{P_0}, R, A \mid Y$.

Assumption 8 *(Conditionally independent misclassification).* Misclassification errors for infection and the post-infection outcome are conditionally independent of one another, or $\tilde{Y} \perp\!\!\!\perp \tilde{S} \mid Y, S$.

We make these assumptions for the remainder of the paper, but we can do parametric sensitivity analyses like those introduced in Section 2.3 to test the sensitivity of our inferences to deviations from these conditions. Let q_{sykzr} be defined as

$$\begin{aligned}q_{sykzr} = {}& \text{sn}_S^s(1-\text{sn}_S)^{1-s}\text{sn}_Y^y(1-\text{sn}_Y)^{1-y}p_{11kjr} + \text{sn}_S^s(1-\text{sn}_S)^{1-s}\text{sp}_Y^{1-y}(1-\text{sp}_Y)^y p_{10kjr} \\ &+ \text{sp}_S^{1-s}(1-\text{sp}_S)^s\text{sp}_Y^{1-y}(1-\text{sp}_Y)^y p_{0*kjr},\end{aligned}$$

recalling the definitions of p_{0*kjr} and p_{1ykjr} in Equation (7) as

$$p_{1ykjr} = \textstyle\sum_{u|u\in\mathcal{S},u_j=1} a_k^u\theta_u^r(\beta_{j,k}^u)^y(1-\beta_{j,k}^u)^{1-y}, \quad p_{0*kjr} = \textstyle\sum_{u|u\in\mathcal{S},u_j=0} a_k^u\theta_u^r. \tag{12}$$

The conditions for the identifiability of the model parameters are outlined in the second major result of this paper, which follows.

Theorem 3. *Let $N_z \geq 2$. Suppose Assumptions 1 to 4 and Assumptions 6 to 8 hold. If both sn_S, sp_S lie in $[0, 1/2)$ or both lie in $(1/2, 1]$, $P_{N_z}(A \mid S^{P_0})$ is at least Kruskal rank $2^{N_z} - 1$ and $P_{N_z}(S^{P_0} \mid R)$ is rank 2^{N_z} then the counterfactual distributions $P(S^{P_0} = u \mid R = r)$, $P(A = k \mid S^{P_0} = u)$ are identifiable as are the quantities $\text{sn}_S, \text{sp}_S, \text{sp}_Y, \text{VE}_{I,jk}^u(k)$, and $\text{VE}_{I,jk}^u$. Furthermore, if sn_Y is unknown (known), distributions $P(Y(z_j) = 1 \mid S^{P_0} = u, A = k)$ are identifiable up to an unknown (known) common constant, $r_Y = \text{sn}_Y + \text{sp}_Y - 1$.*

Theorem 3 allows for a more realistic model of infection measurement than Hudgens & Halloran (2006) and does not require any restrictions on the space of principal strata. The primary benefit of an unrestricted principal strata distribution is that we can jointly infer vaccine efficacy against infection and vaccine efficacy against a post-infection outcome. This will aid in designing comprehensive randomized trials for vaccine efficacy.

The proof of Theorem 3, shown in the Supplementary Materials, relies on a further extension of Kruskal (1977)'s Three-way Array Decomposition Uniqueness theorem. The identifiability results in Theorem 3 suggest the following transparent parameterization: $(\beta_{j,k}^u, \text{sp}_Y, \text{sn}_Y) \to (\tilde{p}_{j,k}^u = (\text{sn}_Y + \text{sp}_Y - 1)\beta_{j,k}^u + (1 - \text{sp}_Y), \text{sp}_Y, \text{sn}_Y)$. The quantities $\tilde{p}_{j,k}^u = P(\tilde{Y} = 1 \mid Z = j, S^{P_0} = u, A = k)$ and sp_Y are identified by the data, while sn_Y is not. This

yields the following asymptotic identification regions for sn_Y and $\beta^u_{j,k}$:

$$\text{sn}_Y \in \left(\max_{u,j,k}(\tilde{p}^u_{j,k}), 1\right), \quad \beta^u_{j,k} \in \left(\frac{\tilde{p}^u_{j,k} - (1 - \text{sp}_Y)}{\text{sp}_Y}, \frac{\tilde{p}^u_{j,k} - (1 - \text{sp}_Y)}{\max_{u,j,k}(\tilde{p}^u_{j,k}) + \text{sp}_Y - 1}\right) \tag{13}$$

This may be useful for policymakers interested in absolute risk of post-infection outcomes to forecast the burden on healthcare centers under different vaccination policies.

We will present a final corollary that will be useful in our applied examples:

COROLLARY 3. *Suppose in addition to Assumptions 1 to 4 and Assumptions 6 to 8, researchers do not directly observe A, but instead observe a misclassified version of A, $\tilde{A}$, such that the following nondifferential error assumption holds: $\tilde{A} \perp\!\!\!\perp \tilde{S}, \tilde{Y}, Y(z,a), R, Z, S^{P_0} \mid A$. If both sn_S, sp_S lie in $[0, 1/2)$ or both lie in $(1/2, 1]$, $P_{N_z}(\tilde{A} \mid S^{P_0})$ is at least Kruskal rank $2^{N_z} - 1$ and $P_{N_z}(S^{P_0} \mid R)$ is rank 2^{N_z} then the counterfactual distributions $P(S^{P_0} = u \mid R = r)$, $P(\tilde{A} = k \mid S^{P_0} = u)$ are identifiable as are the quantities $\text{sn}_S, \text{sp}_S, \text{sp}_Y$, and $\text{VE}^u_{I,jk}$. Furthermore, if sn_Y is unknown (known), distributions $P(Y(z_j) = 1 \mid S^{P_0} = u, \tilde{A} = k)$ are identifiable up to an unknown (known) common constant, $r_Y = \text{sn}_Y + \text{sp}_Y - 1$.*

The proof, shown in the Supplementary Materials, follows directly from the proof of Theorem 3 and the nondifferential misclassification error assumption for A.

4. DESIGN OF VACCINE EFFICACY STUDIES

There are several real-world applications for Theorem 3 in vaccine efficacy studies. The first is for quantifying vaccine efficacy against post-infection outcomes like severe illness, medically-attended illness or death, which is the primary motivation for the methods we have developed here. A second is to quantify the impact on vaccination on transmission, as suggested in VanderWeele & Tchetgen Tchetgen (2011).

4.1. Vaccine efficacy against severe symptoms trial design

To show how our model can be used to design a vaccine efficacy study, we consider determining the sample size for two hypothetical clinical trials: one three-arm trial inspired by Monto et al. (2009), and a two-arm trial inspired by Polack et al. (2020). Monto et al. (2009) investigated vaccine efficacy against symptomatic influenza infection in a three-arm, double-blind placebo-controlled randomized trial. Polack et al. (2020) presented the results of the COVID-19 Pfizer vaccination trial, which measured vaccine efficacy against symptomatic infection using a two-arm double-blind placebo-controlled randomized trial. For both trials, we will target a power of 0.8 against an alternative hypothesis that the vaccine efficacy against symptoms is equal to 0.6 for the always-infected stratum (i.e. $S^{P_0} = (1,1,1)$ and $S^{P_0} = (1,1)$). All trials are designed so as to jointly test the efficacy against infection and the efficacy against severe symptoms for the always-infected group.

In order to design our hypothetical trials, we simulate 100 datasets under the alternative hypothesis for each sample size and measure the proportion of datasets in which we reject the null hypothesis. We reject the null when the posterior probability is 0.90 or larger that vaccine efficacy against severe illness is above 0.1 and that the vaccine efficacy against infection is greater than 0.3. We can write the rejection region for Data = $\{(\tilde{S}_i, \tilde{Y}_i, Z_i, R_i, A_i, X_i), 1 \le i \le n\}$ as $\{\text{Data} : P(\text{VE}^{(1,1,1)}_{I,31} > 0.1, \text{VE}_{S,31} > 0.3 \mid \text{Data}) \ge 0.9\}$ for the three-arm trial and $\{\text{Data} : P(\text{VE}^{(1,1)}_{I,21} > 0.1, \text{VE}_{S,21} > 0.3 \mid \text{Data}) \ge 0.9\}$ for the two-arm trial. This decision criterion is akin to that used in Polack et al. (2020), but we

chose 0.9 so as to control the Type 1 error for a null hypothesis of no vaccine efficacy against severe illness.

We use the model defined in Section 3 along with the parametric model in Section 2.3; the computational details are discussed in the Supplementary materials. In each trial we examine two scenarios: one in which we observe the covariate $\tilde{A}$, or A with error, and one in which we observe A directly. Given the results of Theorem 3, we can determine the number of study sites and the number of levels for A that need to be observed in order to point identify the causal estimand of interest. For the three-arm trial, we need at least 8 study sites and a covariate with at least 7 levels, while for the two-arm trial we need only 4 study sites and a covariate with at least 3 levels.

The power calculations are presented in Table 1, which shows power as a function of the sample size, the number of treatments, and whether A or $\tilde{A}$ was measured.

Table 1. *Power against the alternative,* $\text{VE}_S \approxeq 0.4$, $\text{VE}_{I,31}^{(1,1,1)} \approxeq 0.6$ *for* $N_z = 3$, *and* $\text{VE}_S \approxeq 0.5$, $\text{VE}_{I,21}^{(1,1)} \approxeq 0.6$ *for* $N_z = 2$ *for sample sizes of* 4,000 *through* 120,000. *Scenarios in which* A *was measured with error denoted by* $\tilde{A}$, A *otherwise.*

Trial	Measurements	4,000	20,000	40,000	80,000	120,000
3-arm	A	NA	NA	0.52	0.90	0.99
	$\tilde{A}$	NA	NA	0.24	0.85	0.96
2-arm	A	0.08	0.59	0.85	0.93	NA
	$\tilde{A}$	0.03	0.50	0.82	0.94	NA

While these results show that one needs large sample sizes to achieve 80% power for the estimands of interest in both scenarios, this is expected because the principal strata $(1,1,1)$ and $(1,1)$ are each only 3.5% of their respective populations in our simulation studies. This highlights the extent to which power calculations for our models are dependent on principal strata proportions. Furthermore, though the sample sizes seem large, the results for 2-arm trials show Polack et al. (2020) was appropriately sized to detect joint vaccine efficacy against infection and severe symptoms if Assumption 3 and Assumption 4 could be satisfied. This highlights the fact that our model can be used to infer post-infection outcome vaccine efficacy from large real-world studies.

4.2. Household vaccination study

Consider 2-person households recruited into a vaccination study to determine the infectiousness effect, as termed in VanderWeele & Tchetgen Tchetgen (2011). In other words, if one person in the pair is infected, what benefit does the other person in the household derive from the vaccination status of the infected individual? VanderWeele & Tchetgen Tchetgen (2011) considers a trial design in which exactly one member of each household is randomized between vaccination and placebo. We consider a trial in which the only source of infection for the non-randomized individual is from the individual randomized to treatment. This might be a good model for households in which one member is home-bound. Then the set of treatments for each household can be mapped to a categorical treatment: $z_1 \equiv (0,0), z_2 \equiv (1,0)$. Let the intermediate outcome $S(z)$ be the infection status of the randomized household member, let the set of principal strata be $\{(0,0),(1,0),(0,1),(1,1)\}$, and let the outcome $Y(z)$ be the infection status of the unvaccinated individual. The estimand of interest, vaccine efficacy against transmission, is the expected difference in outcome for the unvaccinated individual when the house-

hold member is unvaccinated vs. when the household member is vaccinated for the set of households in the stratum $(1,1)$:

$$\mathrm{VE}_{T,21}^{(1,1)} = P(Y(z_1) = 1 | S^{P_0} = (1,1)) - P(Y(z_2) = 1 | S^{P_0} = (1,1)).$$

VanderWeele & Tchetgen Tchetgen (2011) derive large-sample bounds for this effect, but we can use our method to identify this quantity. Theorem 3 shows that in order to identify this estimand under noisy infection measurements using our method, one would need at least four study sites, a relevant categorical covariate with three levels, and the sensitivity and specificity to both lie in the same half-interval of $[0,1]$. We write the rejection region for Data $= \{(\tilde{S}_i, \tilde{Y}_i, Z_i, R_i, A_i, X_i),\ 1 \le i \le n\}$ as $\{\text{Data} : P(\mathrm{VE}_{T,21}^{(1,1)} > 0, \mathrm{VE}_{S,21} > 0.3 \mid \text{Data}) \ge 0.95\}$ for the two-arm trial. The 0.95 cutoff was chosen to control the Type 1 error rate, as shown in the Supplementary Material.

Table 2. *Power against the alternative that* $\mathrm{VE}_S \approxeq 0.5$, $\mathrm{VE}_{T,21}^{(1,1)} \approxeq 0.16$. *for sample sizes of* 4,000 *to* 80,000. *Scenarios in which* A *was measured with error denoted by* $\tilde{A}$, A *otherwise.*

Measurements	4,000	20,000	40,000	80,000
A	0.17	0.60	0.75	0.95
$\tilde{A}$	0.07	0.61	0.81	0.94

In this example, the sample size should be understood in terms of households, rather than participants. Our method is applicable to scenarios involving partial interference, which in this case is the assumption that treatment statuses of households do not impact one another.

5. Discussion

Policymakers and public health experts can use vaccine efficacy for post-infection outcomes to design more precise vaccination programs. Our method makes inferring these causal estimands feasible in real-world trials where outcomes are measured with error and vaccines cannot be assumed to have a nonnegative effect on infection on every individual. The power of our method is in its generality, as it can be used for any number of treatments, and for any post-infection outcome distribution, although we focus on binary post-infection outcomes in this manuscript. Accordingly, when paired with a parametric likelihood, our method may be more statistically efficient than models identified by likelihood assumptions alone, like that of Zhang et al. (2009). Furthermore, our identifiability results are nonparametric, though we use parametric Bayesian estimators in our examples. If an appropriate covariate cannot be measured, one can use the asymptotic bounds derived in Section 2.5. Furthermore, one can use these methods to design clinical trials, as we show in Section 4. More work is needed to further generalize the procedure to categorical intermediate outcomes, which would allow for more general vaccine efficacy against transmission study designs (VanderWeele & Tchetgen Tchetgen, 2011), as well as applications beyond vaccine efficacy to noncompliance in multi-arm trials where the exclusion restriction could be violated (Cheng & Small, 2006).

A. Outline of the Supplementary Material

We define our notation for principal stratification in vaccine efficacy (VE) in section A. In section F, we give general properties of the Kruskal rank, and extensions to Kruskal (1977) theorems that we derived. We apply these extensions in the context of principle stratification for VE in section B. The proofs of our main results, Theorem 1 and Theorem 3, are given in section D. These proofs are based on results in Appendix F and Appendix B.

A. Notation and definitions

Let Z be the N_z-category discrete variable taking values in the set $\{1,\ldots,N_z\}$ representing treatment. The principal stratum, S^{P_0}, takes values in the set $\{0,1\}^{N_z}$, of which there are 2^{N_z} elements. Let $\mathcal{S}$ be the set of principal strata, which is equal to $\{0,1\}^{N_z}$ when there are no monotonicity assumptions; let $u \in \mathcal{S}$. Recall the Definition 3, which defines the operator ϖ_m as $\varpi_m(\cdot): j \to \{0,1\}^m, j \in \mathbb{N}, j \leq 2^m - 1$ so that $\varpi_m(j)$ is the base-2 representation of j with m digits in $\mathbb{R}^m$. Let the set of treatments be $\{z_1,\ldots,z_{N_z}\}$, with $z \in \{z_1,\ldots,z_{N_z}\}$ The i^{th} element of the vector is represented as $\varpi_m(j)_i$. For example, $\varpi_3(4) = (0,0,1)$ and $\varpi_5(4) = (0,0,1,0,0)$ with $\varpi_5(4)_3 = 1$ The operator's inverse is represented as $\varpi_m(\cdot)^{-1}: \{0,1\}^m \to j, j \leq 2^m - 1$, so $\varpi_m(u)^{-1}$ is the base-10 representation of the binary m-vector u. For example, $\varpi_3((0,0,1))^{-1} = \varpi_5((0,0,1,0,0))^{-1} = 4$. Let A have N_a levels and take values in the set $\{1,\ldots,N_a\}$. Let $P(A \mid R)$ be the $N_a \times N_r$ matrix with $(i,j)^{\text{th}}$ element equal to $P(A = i \mid R = j)$. Let $P_{N_z}(A \mid S^{P_0})$ be the $N_a \times 2^{N_z}$ matrix with $(i,j)^{\text{th}}$ element equal to $P(A = i \mid S^{P_0} = \varpi_{N_z}(j-1))$, and let $P_{N_z}(S^{P_0} \mid R)$ be the $2^{N_z} \times N_r$ matrix with $(i,j)^{\text{th}}$ element equal to $P(S^{P_0} = \varpi_{N_z}(i-1) \mid R = j)$. Let $P(y \mid R, Z = z)$ be the $1 \times R$ matrix with element $(1,j)^{\text{th}}$ equal to $P(y \mid R = j, Z = z)$, and similarly let $P_{N_z}(y \mid S^{P_0}, Z = z)$ be the 1×2^{N_z} matrix with element $(1,j)^{\text{th}}$ equal to $P(y \mid S^{P_0} = \varpi_{N_z}(j-1), Z = z)$. Let $P(y \mid R, Z = z, A = k)$ be the $1 \times R$ matrix with $(1,j)^{\text{th}}$ element equal to $P(y \mid R = j, Z = z, A = k)$, and similarly let $P_{N_z}(y \mid S^{P_0}, Z = z, A = k)$ be the 1×2^{N_z} matrix with element $(1,j)^{\text{th}}$ equal to $P(y \mid S^{P_0} = \varpi_{N_z}(j-1), Z = z, A = k)$. Let the matrix $P_{N_z}(S \mid Z, S^{P_0})$ be in $\mathbb{R}^{2N_z \times 2^{N_z}}$ where column denotes principal stratum $S^{P_0} = \varpi_{N_z}(j-1)$ and row represents a combination $(s,z) \in \{(1,1),(1,2),\ldots,(1,N_z),(0,1),\ldots,(0,N_z)\}$, with $(i,j)^{\text{th}}$ element denoted $P_{N_z}(S \mid Z, S^{P_0})_{ij}$ defined as

$$P_{N_z}(S \mid Z, S^{P_0})_{ij} = \varpi_{N_z}(j-1)_i \mathbb{1}(i \leq N_z) + (1 - \varpi_{N_z}(j-1)_{i-N_z}) \mathbb{1}(i > N_z),$$

and let $P_{N_z}(\tilde{S} \mid Z, S^{P_0})$ be in $\mathbb{R}^{2N_z \times 2^{N_z}}$ with $(i,j)^{\text{th}}$ element denoted $P_{N_z}(\tilde{S} \mid Z, S^{P_0})_{ij}$ defined:

$$\begin{aligned} P_{N_z}(\tilde{S} \mid Z, S^{P_0})_{ij} = {} & \mathrm{sn}_S^{\varpi_{N_z}(j-1)_i} (1 - \mathrm{sp}_S)^{1 - \varpi_{N_z}(j-1)_i} \mathbb{1}(i \leq N_z) \\ & + (1 - \mathrm{sn}_S)^{\varpi_{N_z}(j-1)_{i-N_z}} \mathrm{sp}_S^{1 - \varpi_{N_z}(j-1)_{i-N_z}} \mathbb{1}(i > N_z). \end{aligned}$$

Let B^+ be the Moore-Penrose inverse of the matrix B, $\mathbf{1}_m$ be the m-vector of 1s, $\mathbf{0}_m$ be the m-vector of 0s, and $\mathbf{I}_m$ be the $m \times m$ dimensional identity matrix.

B. Kruskal rank properties related to VE

In this section, we show that (a) the Kruskal rank of the matrix $P_{N_z}(S \mid Z, S^{P_0})$ is 3 for $N_z \geq 2$, (b) the Kruskal rank of the matrix $P_{N_z}(\tilde{S} \mid Z, S^{P_0})$ is 3 for $N_z \geq 2$ when $\mathrm{sn}_S + \mathrm{sp}_S \neq 1$ and (c) the column domains of $P_{N_z}(\tilde{S} \mid Z, S^{P_0})$ are not invariant to column permutation when $\mathrm{sn}_S, \mathrm{sp}_S > 0.5$ or $\mathrm{sn}_S, \mathrm{sp}_S < 0.5$ for $N_z \geq 2$.

LEMMA A1 (KRUSKAL RANK). *The Kruskal rank of the matrix* $P_2(S \mid Z, S^{P_0})$ *is* 3.

$$P_2(S \mid Z, S^{P_0}) = \begin{array}{c} \begin{matrix} (0,0) & (1,0) & (0,1) & (1,1) \end{matrix} \\ \begin{bmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \end{bmatrix} \end{array} \begin{matrix} (s=1, z=1) \\ (s=1, z=2) \\ (s=0, z=1) \\ (s=0, z=2) \end{matrix} \tag{A1}$$

Proof. All subsets of 3 columns of the matrix $P_2(S \mid Z, S^{P_0})$ are of the form:

$$\begin{bmatrix} a & c & e \\ b & d & f \\ 1-a & 1-c & 1-e \\ 1-b & 1-d & 1-f \end{bmatrix}. \tag{A2}$$

Sub-matrices $[a,b]^T, [c,d]^T, [e,f]^T$ are any three (without replacement) columns of the matrix:

$$\begin{bmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}, \tag{A3}$$

which corresponds to the first 2 rows of $P_2(S \mid Z, S^{P_0})$. Matrices (A2) have 4 associated 3×3 minors equal to $a(d-f) - c(b-f) + e(b-d)$. For each of the 4 combinations of 3 columns from eq. (A3), all 4 minors are nonzero. By the determinantal definition of rank, each matrix is rank-3. In contrast, the determinant of $P_2(S \mid Z, S^{P_0})$ is 0. Thus, by Definition 4, the matrix $P_2(S \mid Z, S^{P_0})$ is Kruskal rank 3. □

LEMMA A2 (KRUSKAL RANK). *The Kruskal rank of the matrix* $P_{N_z}(S \mid Z, S^{P_0})$ *is* 3 *for* $N_z \geq 2$.

Proof. For general $N_z = m$ recall that the j-th column of $P_m(S \mid Z, S^{P_0})$ is of the form

$$\begin{bmatrix} \varpi_m(j-1) \\ \mathbf{1}_m - \varpi_m(j-1) \end{bmatrix} \tag{A4}$$

We will proceed using induction: We have shown in Lemma A1 that the Kruskal rank of $P_2(S \mid Z, S^{P_0})$ when $N_z = 2$ is 3. Suppose that the Kruskal rank of the $2n \times 2^n$ matrix $P_n(S \mid Z, S^{P_0})$ for $N_z = n$ is 3. Let $N_z = n+1$ so that the matrix $P_{n+1}(S \mid Z, S^{P_0})$ is in $\mathbb{R}^{2(n+1)\times 2^{n+1}}$. The j-th column of $P_{n+1}(S \mid Z, S^{P_0})$ is

$$\begin{bmatrix} \varpi_n(j-1) \\ 0 \\ \mathbf{1}_n - \varpi_n(j-1) \\ 1 \end{bmatrix} \tag{A5}$$

for $j \in \{1, \ldots, 2^n\}$ and

$$\begin{bmatrix} \varpi_n(j-2^n-1) \\ 1 \\ \mathbf{1}_n - \varpi_n(j-2^n-1) \\ 0 \end{bmatrix} \tag{A6}$$

for $j \in \{2^n+1,\dots,2^{n+1}\}$. The sets of columns of 3 have several configurations. For any 3-column submatrix constructed from columns $j,k,m \in \{1,\dots,2^n\}$:

$$\operatorname{rank}\left(\begin{bmatrix} \varpi_n(j-1) & \varpi_n(k-1) & \varpi_n(m-1) \\ 0 & 0 & 0 \\ \mathbf{1}_n - \varpi_n(j-1) & \mathbf{1}_n - \varpi_n(k-1) & \mathbf{1}_n - \varpi_n(m-1) \\ 1 & 1 & 1 \end{bmatrix}\right) = 3 \tag{A7}$$

by the induction hypothesis. The submatrices constructed from columns $j,k,m \in \{2^n+1,\dots,2^{n+1}\}$ are also rank 3 by the induction hypothesis. For the 3-column submatrix constructed from columns $j,k \in \{1,\dots,2^n\}$ and $m \in \{2^n+1,\dots,2^{n+1}\}$:

$$\operatorname{rank}\left(\begin{bmatrix} \varpi_n(j-1) & \varpi_n(k-1) & \varpi_n(m-2^n-1) \\ 0 & 0 & 1 \\ \mathbf{1}_n - \varpi_n(j-1) & \mathbf{1}_n - \varpi_n(k-1) & \mathbf{1}_n - \varpi_n(m-2^n-1) \\ 1 & 1 & 0 \end{bmatrix}\right) \tag{A8}$$

$$\geq \operatorname{rank}\left(\begin{bmatrix} \varpi_n(j-1) & \varpi_n(k-1) \\ \mathbf{1}_n - \varpi_n(j-1) & \mathbf{1}_n - \varpi_n(k-1) \\ 1 & 1 \end{bmatrix}\right) + \operatorname{rank}(1) = 3 \tag{A9}$$

where the inequality follows from Lemma A13 and the induction hypothesis. As the rank of the submatrix is ≤ 3, it follows the rank is 3. The ranks of submatrices with $j \in \{1,\dots,2^n\}$ and $k,m \in \{2^n+1,\dots,2^{n+1}\}$ follow similarly. It follows that all 3-column submatrices of $P_n(S \mid Z, S^{P_0})$ are rank 3. By Definition 4, $P_n(S \mid Z, S^{P_0})$ is Kruskal rank 3. □

LEMMA A3 (KRUSKAL RANK $P_2(\tilde{S} \mid Z, S^{P_0})$). *The Kruskal rank of*

$$\begin{array}{c} \begin{matrix} (0,0) & (1,0) & (0,1) & (1,1) \end{matrix} \\ \begin{bmatrix} 1-\mathrm{sp}_S & \mathrm{sn}_S & 1-\mathrm{sp}_S & \mathrm{sn}_S \\ 1-\mathrm{sp}_S & 1-\mathrm{sp}_S & \mathrm{sn}_S & \mathrm{sn}_S \\ \mathrm{sp}_S & 1-\mathrm{sn}_S & \mathrm{sp}_S & 1-\mathrm{sn}_S \\ \mathrm{sp}_S & \mathrm{sp}_S & 1-\mathrm{sn}_S & 1-\mathrm{sn}_S \end{bmatrix} \end{array} \begin{array}{l} (s=1,z=1) \\ (s=1,z=2) \\ (s=0,z=1) \\ (s=0,z=2) \end{array} \tag{A10}$$

is 3 *as long as* $\mathrm{sn}_S + \mathrm{sp}_S \neq 1$.

Proof. All 3 column submatrices of matrix (A7) are of the same form as matrix (A2). By the same logic as set forth in Lemma A1, these submatrices have a common maximal minor of

$$a(d-f) - c(b-f) + e(b-d).$$

The quantities a,b,c,d,e,f are the elements of the 2×3 matrix

$$\begin{bmatrix} a & c & e \\ b & d & f \end{bmatrix} \tag{A11}$$

in which $(a,b)^T, (c,d)^T, (e,f)^T$ are any 3 columns drawn without replacement from the 2×4 submatrix of Equation (A7):

$$\begin{bmatrix} 1-\mathrm{sp}_S & \mathrm{sn}_S & 1-\mathrm{sp}_S & \mathrm{sn}_S \\ 1-\mathrm{sp}_S & 1-\mathrm{sp}_S & \mathrm{sn}_S & \mathrm{sn}_S \end{bmatrix}. \tag{A12}$$

These minors are all equal to (up to a factor of –1):

$$(1-\mathrm{sn}_S-\mathrm{sp}_S)^2,$$

which can be seen after a brute-force calculation. The minors are nonzero for all $\mathrm{sn}_S, \mathrm{sp}_S \in [0,1]$ such that $\mathrm{sn}_S + \mathrm{sp}_S \neq 1$. Thus, by the determinantal rank definition, all 3 column matrices are

rank 3. In contrast, the determinant of $P_2(\tilde{S} \mid Z, S^{P_0})$ is 0 for all values of $\mathrm{sn}_S, \mathrm{sp}_S$. Thus by the definition of Kruskal rank in Definition 4, $k_{P_2(\tilde{S}|Z,S^{P_0})} = 3$. □

LEMMA A4 (KRUSKAL RANK $P_{N_z}(\tilde{S} \mid Z, S^{P_0})$, $N_z \geq 2$). *The Kruskal rank of* $P(\tilde{S} \mid Z, S^{P_0})$ *for* $N_z \geq 2$ *is* 3 *as long as* $\mathrm{sn}_S + \mathrm{sp}_S \neq 1$.

Proof. We proceed by induction. For $N_z = 2$, Lemma A3 shows that the Kruskal rank is 3. Let $N_z = n$ for $n > 2$. Recall that $P_n(\tilde{S} \mid Z, S^{P_0})$ is the $2n \times 2^n$ matrix with column j

$$\begin{bmatrix} s_j \\ \mathbf{1}_n - s_j \end{bmatrix}$$

with the i^{th} element of s_j denoted s_{ij} and defined as:

$$s_{ij} = \mathrm{sn}_S^{\varpi_n(j-1)_i}(1 - \mathrm{sp}_S)^{1-\varpi_n(j-1)_i}.$$

The induction hypothesis is that the Kruskal rank of $P_n(\tilde{S} \mid Z, S^{P_0})$ is 3. The columns of $P_{n+1}(\tilde{S} \mid Z, S^{P_0})$ are of the form

$$\begin{bmatrix} s_j \\ 1 - \mathrm{sp}_S \\ \mathbf{1}_n - s_j \\ \mathrm{sp}_S \end{bmatrix}$$

for $j \in \{1, \ldots, 2^n\}$, and

$$\begin{bmatrix} s_{j-2^n} \\ \mathrm{sn}_S \\ \mathbf{1}_n - s_{j-2^n} \\ 1 - \mathrm{sn}_S \end{bmatrix}$$

for $j \in \{2^n + 1, \ldots, 2^{n+1}\}$. The 3-column submatrices of $P_{N_z}(\tilde{S} \mid Z, S^{P_0})$ made from column j, ℓ, m indices fall into several classes. When $j, \ell, m \in \{1, \ldots, 2^n\}$, $j, \ell, m \in \{2^n + 1, \ldots, 2^{n+1}\}$ or $j, \ell \in \{1, \ldots, 2^n\}, m \in \{2^n + 1, \ldots, 2^{n+1}\} \setminus \{j + 2^n, \ell + 2^n\}$, $j \in \{1, \ldots, 2^n\}, m, \ell \in \{2^n + 1, \ldots, 2^{n+1}\} \setminus \{j + 2^n\}$ all matrices are rank 3 by the induction hypothesis. When $j, \ell \in \{1, \ldots, 2^n\}$ but $m \in \{j + 2^n, \ell + 2^n\}$ the submatrix is

$$\begin{bmatrix} s_j & s_\ell & s_{m-2^n} \\ 1 - \mathrm{sp}_S & 1 - \mathrm{sp}_S & \mathrm{sn}_S \\ \mathbf{1}_n - s_j & \mathbf{1}_n - s_\ell & \mathbf{1}_n - s_{m-2^n} \\ \mathrm{sp}_S & \mathrm{sp}_S & 1 - \mathrm{sn}_S \end{bmatrix}.$$

WLOG, let $m = j + 2^n$. This leads to the submatrix:

$$\begin{bmatrix} s_j & s_\ell & s_j \\ 1 - \mathrm{sp}_S & 1 - \mathrm{sp}_S & \mathrm{sn}_S \\ \mathbf{1}_n - s_j & \mathbf{1}_n - s_\ell & \mathbf{1}_n - s_j \\ \mathrm{sp}_S & \mathrm{sp}_S & 1 - \mathrm{sn}_S \end{bmatrix}.$$

The rank of this submatrix is

$$\begin{aligned}
\operatorname{rank}\begin{bmatrix} s_j & s_\ell & s_j \\ 1-\mathrm{sp}_S & 1-\mathrm{sp}_S & \mathrm{sn}_S \\ \mathbf{1}_n - s_j & \mathbf{1}_n - s_\ell & \mathbf{1}_n - s_j \\ \mathrm{sp}_S & \mathrm{sp}_S & 1-\mathrm{sn}_S \end{bmatrix} &= \operatorname{rank}\begin{bmatrix} s_j & s_\ell & s_j \\ 1-\mathrm{sp}_S & 1-\mathrm{sp}_S & \mathrm{sn}_S \\ \mathbf{1}_n - s_j & \mathbf{1}_n - s_\ell & \mathbf{1}_n - s_j \\ 1 & 1 & 1 \end{bmatrix} \\
&= \operatorname{rank}\begin{bmatrix} s_j & s_\ell & s_j \\ \mathbf{1}_n - s_j & \mathbf{1}_n - s_\ell & \mathbf{1}_n - s_j \\ 1-\mathrm{sp}_S & 1-\mathrm{sp}_S & \mathrm{sn}_S \\ 0 & 0 & 1-\frac{\mathrm{sn}_S}{1-\mathrm{sp}_S} \end{bmatrix} \\
&\geq \operatorname{rank}\begin{bmatrix} s_j & s_\ell \\ \mathbf{1}_n - s_j & \mathbf{1}_n - s_\ell \\ 1-\mathrm{sp}_S & 1-\mathrm{sp}_S \\ 0 & 0 \end{bmatrix} + \operatorname{rank}(1-\frac{\mathrm{sn}_S}{1-\mathrm{sp}_S}) \\
&= 3.
\end{aligned}$$

The inequality follows from Lemma A13. Other scenarios follow similarly. □

Lemma A5 (Domain restriction lemma). *If* $\mathrm{sn}_S, \mathrm{sp}_S \in [0, 0.5)$ *or* $\mathrm{sn}_S, \mathrm{sp}_S \in (0.5, 1]$, *the matrix* $P_{N_z}(\tilde{S} \mid Z, S^{P_0}) \in \mathbb{R}^{2N_z \times 2^{N_z}}$ *has column domains that are not invariant to column permutation.*

Proof. We prove Lemma A5 by induction on N_z. The base case is $N_z = 2$. Let P be a 4×4 permutation matrix and let $P_2(\tilde{S} \mid Z, S^{P_0})$ be

$$\begin{array}{c} \begin{matrix} (0,0) & (1,0) & (0,1) & (1,1) \end{matrix} \\ \begin{bmatrix} 1-\mathrm{sp}_S & \mathrm{sn}_S & 1-\mathrm{sp}_S & \mathrm{sn}_S \\ 1-\mathrm{sp}_S & 1-\mathrm{sp}_S & \mathrm{sn}_S & \mathrm{sn}_S \\ \mathrm{sp}_S & 1-\mathrm{sn}_S & \mathrm{sp}_S & 1-\mathrm{sn}_S \\ \mathrm{sp}_S & \mathrm{sp}_S & 1-\mathrm{sn}_S & 1-\mathrm{sn}_S \end{bmatrix} \end{array} \begin{array}{l} (s=1, z=1) \\ (s=1, z=2) \\ (s=0, z=1) \\ (s=0, z=2) \end{array} \tag{A13}$$

Recall from the definition in Appendix A that the column indices $\{1,2,3,4\}$ of $P_2(\tilde{S} \mid Z, S^{P_0})$ map to the following principal strata S^{P_0}: $\varpi_2(0), \varpi_2(1), \varpi_2(2), \varpi_2(3)$. In other words, column index j is mapped to S^{P_0} via the relation $\varpi_2(j-1)$. We consider permutation matrix P without loss of generality, and other cases are similarly shown,

$$P = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

Let $\mathcal{C} = [0,1]$, and let $\mathcal{A}$ be one of two half intervals of $[0,1]$: $[0, 0.5)$ or $(0.5, 1]$. Let $\mathcal{B} = \mathcal{C} \setminus \mathcal{A}$. Note that $P_2(\tilde{S} \mid Z, S^{P_0})$ maps $\mathcal{C} \times \mathcal{C}$ to a matrix with elements in $\mathcal{C}$. Let $1 - \mathrm{sp}_S \in \mathcal{A}$ and let $\mathrm{sn}_S \in \mathcal{B}$ and suppose that the column domains for $P_2(\tilde{S} \mid Z, S^{P_0})$ are not invariant after permutation by matrix P. Then we have the following domain for the map given by $P_2(\tilde{S} \mid Z, S^{P_0})$:

$$P_2(\tilde{S} \mid Z, S^{P_0})\,|_{\mathcal{A}\times\mathcal{B}} \colon \mathcal{A} \times \mathcal{B} \to \begin{array}{c} \begin{matrix} (0,0) & (1,0) & (0,1) & (1,1) \end{matrix} \\ \begin{bmatrix} \mathcal{A} & \mathcal{A} & \mathcal{A} & \mathcal{A} \\ \mathcal{A} & \mathcal{A} & \mathcal{A} & \mathcal{A} \\ \mathcal{B} & \mathcal{B} & \mathcal{B} & \mathcal{B} \\ \mathcal{B} & \mathcal{B} & \mathcal{B} & \mathcal{B}) \end{bmatrix} \end{array}$$

However, we have,

$$\bar{P}_2(\tilde{S} \mid Z, S^{P_0}) \mid_{\mathcal{A}\times\mathcal{B}} = P_2(\tilde{S} \mid Z, S^{P_0}) \mid_{\mathcal{A}\times\mathcal{B}} P \tag{A14}$$

$$= \begin{matrix} \\ 1-\mathrm{sp}_S \\ 1-\mathrm{sp}_S \\ \mathrm{sp}_S \\ \mathrm{sp}_S \end{matrix} \begin{matrix} (0,0) \\ \mathrm{sn}_S \\ 1-\mathrm{sn}_S \\ \mathrm{sp}_S \\ \mathrm{sp}_S \end{matrix} \begin{matrix} (1,0) \\ 1-\mathrm{sp}_S \\ 1-\mathrm{sp}_S \\ \mathrm{sn}_S \\ 1-\mathrm{sn}_S \end{matrix} \begin{matrix} (0,1) \\ \mathrm{sn}_S \\ \mathrm{sn}_S \\ 1-\mathrm{sp}_S \\ 1-\mathrm{sn}_S \end{matrix} \begin{matrix} (1,1) \\ \mathrm{sn}_S \\ \mathrm{sn}_S \\ 1-\mathrm{sn}_S \\ 1-\mathrm{sn}_S \end{matrix} \begin{bmatrix} 0&0&0&1 \\ 0&1&0&0 \\ 0&0&1&0 \\ 1&0&0&0 \end{bmatrix} \tag{A15}$$

$$= \begin{matrix} \\ \mathrm{sn}_S \\ \mathrm{sn}_S \\ 1-\mathrm{sn}_S \\ 1-\mathrm{sn}_S \end{matrix} \begin{matrix} (0,0) \\ 1-\mathrm{sp}_S \\ 1-\mathrm{sp}_S \\ \mathrm{sp}_S \\ \mathrm{sp}_S \end{matrix} \begin{matrix} (1,0) \\ \mathrm{sn}_S \\ 1-\mathrm{sp}_S \\ 1-\mathrm{sn}_S \\ \mathrm{sp}_S \end{matrix} \begin{matrix} (0,1) \\ 1-\mathrm{sp}_S \\ \mathrm{sn}_S \\ \mathrm{sp}_S \\ 1-\mathrm{sn}_S \end{matrix} \begin{matrix} (1,1) \\ \mathrm{sn}_S \\ \mathrm{sn}_S \\ \mathrm{sp}_S \\ \mathrm{sp}_S \end{matrix} \tag{A16}$$

But we see that the column domains are invariant after column permutation:

$$\bar{P}_2(\tilde{S} \mid Z, S^{P_0}) \mid_{\mathcal{A}\times\mathcal{B}} : \mathcal{A}\times\mathcal{B} \to \begin{matrix} (0,0) & (1,0) & (0,1) & (1,1) & \\ \mathcal{A} & \mathcal{A} & \mathcal{A} & \mathcal{A} & (s=1,z=1) \\ \mathcal{A} & \mathcal{A} & \mathcal{A} & \mathcal{A} & (s=1,z=2) \\ \mathcal{B} & \mathcal{B} & \mathcal{B} & \mathcal{B} & (s=0,z=1) \\ \mathcal{B} & \mathcal{B} & \mathcal{B} & \mathcal{B} & (s=0,z=2) \end{matrix}$$

In order for the columns $\bar{P}_2(\tilde{S} \mid Z, S^{P_0}) \mid_{\mathcal{A}\times\mathcal{B}}$ to be on the same domain as $P_2(\tilde{S} \mid Z, S^{P_0}) \mid_{\mathcal{A}\times\mathcal{B}}$, a necessary and sufficient condition is that sn_S and $1-\mathrm{sp}_S$ are on the same domain. In other words, $\{\mathrm{sn}_S \in \mathcal{A}, \mathrm{sp}_S \in \mathcal{B}\}$ or $\{\mathrm{sn}_S \in \mathcal{B}, \mathrm{sp}_S \in \mathcal{A}\}$.

Thus $\bar{P}_2(\tilde{S} \mid Z, S^{P_0}) \mid_{\mathcal{A}\times\mathcal{B}}$ maps $(\mathrm{sn}_S, \mathrm{sp}_S)$ to the same space that $P_2(\tilde{S} \mid Z, S^{P_0}) \mid_{\mathcal{A}\times\mathcal{B}}$. We contradict our statement that the columns are not invariant to permutation.

The case for $N_z > 2$. Let $N_z = n > 2$ and let the column domains of $P_n(\tilde{S} \mid Z, S^{P_0})$ be not invariant to permutation. Furthermore suppose that $\mathrm{sn}_S, {}_S \in \mathcal{A}$ or $\mathrm{sn}_S, {}_S \in \mathcal{B}$. Then matrix $P_{n+1}(\tilde{S} \mid Z, S^{P_0})$ has columns

$$\begin{bmatrix} s_j \\ 1-\mathrm{sp}_S \\ \mathbf{1}_n - s_j \\ \mathrm{sp}_S \end{bmatrix}$$

for $j \in \{1, \ldots, 2^n\}$ and

$$\begin{bmatrix} s_{j-2^n} \\ \mathrm{sn}_S \\ \mathbf{1}_n - s_{j-2^n} \\ 1-\mathrm{sn}_S \end{bmatrix}$$

for $j \in \{2^n+1, \ldots, 2^{n+1}\}$. Permuting any two columns $j, k \in \{1, \ldots, 2^n\}$ or $j, k \in \{2^n+1, \ldots, 2^{n+1}\}$ yields different column domains given the induction hypothesis. If $j \in \{1, \ldots, 2^n\}$ and $k = j + 2^n$, then the columns are

$$\begin{bmatrix} s_j & s_j \\ 1-\mathrm{sp}_S & \mathrm{sn}_S \\ \mathbf{1}_n - s_j & \mathbf{1}_n - s_j \\ \mathrm{sp}_S & 1-\mathrm{sn}_S \end{bmatrix}$$

Let the domain of s_j be $\mathcal{D}$, and let $\mathcal{D}^{\mathsf{c}} = [0,1]^n \setminus \mathcal{D}$ be the domain of $\mathbf{1}_n - s_j$. Then the domains are

$$\begin{bmatrix} \mathcal{D} & \mathcal{D} \\ \mathcal{A} & \mathcal{B} \\ \mathcal{D}^{\mathsf{c}} & \mathcal{D}^{\mathsf{c}} \\ \mathcal{B} & \mathcal{A} \end{bmatrix}$$

if $\mathrm{sp}_S, \mathrm{sn}_S \in \mathcal{B}$ and

$$\begin{bmatrix} \mathcal{D} & \mathcal{D} \\ \mathcal{B} & \mathcal{A} \\ \mathcal{D}^{\mathsf{c}} & \mathcal{D}^{\mathsf{c}} \\ \mathcal{A} & \mathcal{B} \end{bmatrix}$$

if $\mathrm{sp}_S, \mathrm{sn}_S \in \mathcal{A}$. These two columns are not invariant to permutation. Because no two columns may be interchanged without a change in domain, right multiplying $P_{n+1}(\tilde{S} \mid Z, S^{P_0})$ by any $2^{n+1} \times 2^{n+1}$ permutation matrix $P \neq \mathbf{I}_{n+1}$ to will yield a matrix with different column domains than $P_{n+1}(\tilde{S} \mid Z, S^{P_0})$. □

C. Rank properties related to VE

In this section we show that when $N_z \geq 2$ (1) the rank of $P_{N_z}(S \mid Z, S^{P_0}) = N_z + 1$, and (2) $P_{N_z}(\tilde{S} \mid Z, S^{P_0}) = N_z + 1$ when $\mathrm{sn}_S + \mathrm{sp}_S \neq 1$.

Lemma A6 (Rank $P_2(S \mid Z, S^{P_0})$). *The rank of matrix $P_2(S \mid Z, S^{P_0})$ is 3.*

Proof.

$$P_2(S \mid Z, S^{P_0}) = \begin{array}{c} \begin{matrix} (0,0) & (1,0) & (0,1) & (1,1) \end{matrix} \\ \begin{bmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \end{bmatrix} \end{array} \begin{matrix} (s=1, z=1) \\ (s=1, z=2) \\ (s=0, z=1) \\ (s=0, z=2) \end{matrix} \tag{A1}$$

The determinant of $P_2(S \mid Z, S^{P_0})$ is zero, but the determinant of the 3-minor $M_{4,4}$ is 1, so by the determinantal definition of rank, the matrix is rank 3. □

Lemma A7 (Rank $P_{N_z}(S \mid Z, S^{P_0})$). *The rank of the matrix $P_{N_z}(S \mid Z, S^{P_0})$ is $N_z + 1$ for $N_z \geq 2$.*

Proof. For general $N_z = m$ recall that the j-th column of $P_m(S \mid Z, S^{P_0})$ is of the form

$$\begin{bmatrix} \varpi_m(j-1) \\ \mathbf{1}_m - \varpi_m(j-1) \end{bmatrix} \tag{A2}$$

We will proceed using induction: We have shown in Lemma A6 that the rank of $P_2(S \mid Z, S^{P_0})$ when $N_z = 2$ is 3. Suppose that the rank of the $2n \times 2^n$ matrix $P_n(S \mid Z, S^{P_0})$ for $N_z = n$ is $n+1$. Let $N_z = n+1$ so that the matrix $P_{n+1}(S \mid Z, S^{P_0})$ is in $\mathbb{R}^{2(n+1) \times 2^{n+1}}$. The j-th column of $P_{n+1}(S \mid Z, S^{P_0})$ is

$$\begin{bmatrix} \varpi_n(j-1) \\ 0 \\ \mathbf{1}_n - \varpi_n(j-1) \\ 1 \end{bmatrix} \tag{A3}$$

for $j \in \{1, \ldots, 2^n\}$ and

$$\begin{bmatrix} \varpi_n(j-2^n-1) \\ 1 \\ \mathbf{1}_n - \varpi_n(j-2^n-1) \\ 0 \end{bmatrix} \tag{A4}$$

for $j \in \{2^n+1, \ldots, 2^{n+1}\}$. After a row permutation we can express $P_{n+1}(S \mid Z, S^{P_0})$ as a block matrix:

$$\begin{bmatrix} P_n(S \mid Z, S^{P_0}) & P_n(S \mid Z, S^{P_0}) \\ \mathbf{0}_{2^n}^T & \mathbf{1}_{2^n}^T \\ \mathbf{1}_{2^n}^T & \mathbf{0}_{2^n}^T \end{bmatrix}$$

Recall that by construction the sum of the i^{th} row with the $(i+n)^{\text{th}}$ row of $P_n(S \mid Z, S^{P_0})$ is $\mathbf{1}_{2^n}^T$ for $i \le n$. Then $\mathbf{1}_{2^n} \in \text{range}(P_n(S \mid Z, S^{P_0})^T)$. This means we can apply Lemma A14. By Lemma A14, $\text{rank}\left(P_{n+1}(S \mid Z, S^{P_0})\right)$ is

$$\text{rank}\left(P_{n+1}(S \mid Z, S^{P_0})\right) = \text{rank}\left(P_n(S \mid Z, S^{P_0})\right) + \text{rank}\left(\begin{bmatrix} \mathbf{1}_{2^n}^T \\ \mathbf{0}_{2^n}^T \end{bmatrix} - \begin{bmatrix} \mathbf{0}_{2^n}^T \\ \mathbf{1}_{2^n}^T \end{bmatrix}\right) \tag{A5}$$

$$= n + 1 + 1. \tag{A6}$$

Lemma A8 (Rank $P_2(\tilde{S} \mid Z, S^{P_0})$). *The rank of*

$$\begin{array}{c} \begin{matrix} (0,0) & (1,0) & (0,1) & (1,1) \end{matrix} \\ \begin{bmatrix} 1-\text{sp}_S & \text{sn}_S & 1-\text{sp}_S & \text{sn}_S \\ 1-\text{sp}_S & 1-\text{sp}_S & \text{sn}_S & \text{sn}_S \\ \text{sp}_S & 1-\text{sn}_S & \text{sp}_S & 1-\text{sn}_S \\ \text{sp}_S & \text{sp}_S & 1-\text{sn}_S & 1-\text{sn}_S \end{bmatrix} \end{array} \begin{matrix} (s=1, z=1) \\ (s=1, z=2) \\ (s=0, z=1) \\ (s=0, z=2) \end{matrix} \tag{A7}$$

is 3 *as long as* $\text{sn}_S + \text{sp}_S \neq 1$.

Proof. The determinant of $P_2(\tilde{S} \mid Z, S^{P_0})$ is 0. The determinant of the 3-minor $M_{4,4}$ is $(1 - \text{sn}_S - \text{sp}_S)^2$ which is nonzero as long as $\text{sn}_S + \text{sp}_S \neq 1$. □

Lemma A9 (Rank $P_{N_z}(\tilde{S} \mid Z, S^{P_0})$, $N_z \ge 2$). *The rank of* $P(\tilde{S} \mid Z, S^{P_0})$ *for* $N_z \ge 2$ *is* $N_z + 1$ *as long as* $\text{sn}_S + \text{sp}_S \neq 1$.

Proof. We proceed by induction. For $N_z = 2$, Lemma A8 shows that the rank is 3. Let $N_z = n$ for $n > 2$. Recall that $P_n(\tilde{S} \mid Z, S^{P_0})$ is the $2n \times 2^n$ matrix with column j

$$\begin{bmatrix} s_j \\ \mathbf{1}_n - s_j \end{bmatrix}$$

with the i^{th} element of s_j denoted s_{ij} and defined as:

$$s_{ij} = \text{sn}_S^{\varpi_n(j-1)_i}(1-\text{sp}_S)^{1-\varpi_n(j-1)_i}.$$

The induction hypothesis is that the rank of $P_n(\tilde{S} \mid Z, S^{P_0})$ is $n+1$. The columns of $P_{n+1}(\tilde{S} \mid Z, S^{P_0})$ are of the form

$$\begin{bmatrix} s_j \\ 1-\text{sp}_S \\ \mathbf{1}_n - s_j \\ \text{sp}_S \end{bmatrix}$$

for $j \in \{1, \ldots, 2^n\}$, and

$$\begin{bmatrix} s_{j-2^n} \\ \text{sn}_S \\ \mathbf{1}_n - s_{j-2^n} \\ 1 - \text{sn}_S \end{bmatrix}$$

for $j \in \{2^n + 1, \ldots, 2^{n+1}\}$. After a row permutation we can express $P_{n+1}(\tilde{S} \mid Z, S^{P_0})$ as a block matrix:

$$\begin{bmatrix} P_n(\tilde{S} \mid Z, S^{P_0}) & P_n(\tilde{S} \mid Z, S^{P_0}) \\ (1 - \text{sp}_S)\mathbf{1}_{2^n}^T & \text{sn}_S \mathbf{1}_{2^n}^T \\ \text{sp}_S \mathbf{1}_{2^n}^T & (1 - \text{sn}_S)\mathbf{1}_{2^n}^T \end{bmatrix}$$

Recall that by construction the sum of the i^{th} row with the $(i+n)^{\text{th}}$ row of $P_n(\tilde{S} \mid Z, S^{P_0})$ is $\mathbf{1}_{2^n}^T$ for $i \leq n$. Then by Lemma A14, $\text{rank}\left(P_{n+1}(S \mid Z, S^{P_0})\right)$ is

$$\text{rank}\left(P_{n+1}(S \mid Z, S^{P_0})\right) = \text{rank}\left(P_n(S \mid Z, S^{P_0})\right) + \text{rank}\left(\begin{bmatrix} \text{sn}_S \mathbf{1}_{2^n}^T \\ (1 - \text{sn}_S)\mathbf{1}_{2^n}^T \end{bmatrix} - \begin{bmatrix} (1 - \text{sp}_S)\mathbf{1}_{2^n}^T \\ \text{sp}_S \mathbf{1}_{2^n}^T \end{bmatrix}\right) \tag{A8}$$

$$= n + 1 + 1 \tag{A9}$$

given that $\text{sn}_S + \text{sp}_S \neq 1$. □

D. Main results

Proof. Proof of Theorem 1
By Assumptions 1 to 2 $P(Y(z_l) = y \mid S^{P_0}, A = k) = P_{N_z}(y \mid S^{P_0}, Z = z_l, A = k)$. The causal estimand VE_I is a functional of $P_{N_z}(y \mid S^{P_0}, Z = z_l, A = k)$ and $P_{N_z}(A \mid S^{P_0})$, so it is sufficient to identify these distributions to show identifiability of the causal estimand. Let X be the three way array with dimensions $2N_z \times N_a \times N_r$ with $(i,k,r)^{\text{th}}$ element $P(S = \mathbb{1}(i \leq N_z), A = k \mid Z = z_{i - N_z \mathbb{1}(z > N_z)}, R = r)$. Recall the definitions in Appendix A of $P_{N_z}(S \mid Z, S^{P_0})$, $P_{N_z}(S^{P_0} \mid R)$, and $P_{N_z}(A \mid S^{P_0})$. Then the observable probabilities can be written as

$$P(S = \mathbb{1}(i \leq N_z), A = k \mid Z = z_{i - N_z \mathbb{1}(i > N_z)}, R = r) = \sum_{j=1}^{2^{N_z}} P_{N_z}(S \mid Z, S^{P_0})_{i,j} P_{N_z}(S^{P_0} \mid R)^T_{r,j} P_{N_z}(A \mid S^{P_0})_{k,j}.$$

By Lemma A2, $k_{P_{N_z}(S|Z,S^{P_0})} = 3$ and by Lemma A7, $\text{rank}(P_{N_z}(S \mid Z, S^{P_0})) = N_z + 1$. Furthermore, $P_{N_z}(S \mid Z, S^{P_0})$ is a fixed, known matrix. This means Lemma A11 is applicable to the decomposition of 3-way array X. By the assumptions stated in Theorem 1, $\text{rank}(P_{N_z}(S^{P_0} \mid R)^T) = 2^{N_z}$ and $P_{N_z}(S^{P_0} \mid R)^T \in \mathbb{R}^{N_r \times 2^{N_z}}$. Then it follows that by Definition 4, $k_{P_{N_z}(S^{P_0}|R)^T} = 2^{N_z}$. Given that $k_{P_{N_z}(A|S^{P_0})} \geq 2^{N_z} - 1$ as stated in Theorem 1, the conditions in Lemma A11 hold:

$$\min(3, 2^{N_z}) + 2^{N_z} - 1 \geq 2^{N_z} + 2 \tag{A1}$$

$$\min(3, 2^{N_z} - 1) + 2^{N_z} \geq 2^{N_z} + 2 \tag{A2}$$

$$\tag{A3}$$

and

$$\text{rank}(P_{N_z}(S \mid Z, S^{P_0})) + \text{rank}(P_{N_z}(S^{P_0} \mid R)) + \text{rank}(P_{N_z}(A \mid S^{P_0})) \tag{A4}$$

$$\geq N_z + 1 + 2^{N_z} + 2^{N_z} - 1 \tag{A5}$$

$$= N_z + 2^{N_z + 1} \tag{A6}$$

by the fact that $\text{rank}(P_{N_z}(A \mid S^{P_0})) \geq k_{(P_{N_z}(A|S^{P_0})}$. Also

$$N_z + 2^{N_z+1} - 2(2^{N_z} - 1) = N_z - 1 \tag{A7}$$

$$\geq \begin{cases} \min(N_z - 2, \text{rank}(P_{N_z}(A \mid S^{P_0})) - k_{(P_{N_z}(A|S^{P_0})} \\ \min(N_z - 2, 0) \end{cases} . \tag{A8}$$

Applying the results from Lemma A11, the triple product decomposition of array X, denoted as $X = [P_{N_z}(S \mid Z, S^{P_0}), P_{N_z}(S^{P_0} \mid R)^T, P_{N_z}(A \mid S^{P_0})]$, is unique. Then it follows that for two sets of matrices $[P_{N_z}(S \mid Z, S^{P_0}), P_{N_z}(S^{P_0} \mid R)^T, P_{N_z}(A \mid S^{P_0})]$, $[P_{N_z}(S \mid Z, S^{P_0})', (P_{N_z}(S^{P_0} \mid R)^T)', P_{N_z}(A \mid S^{P_0})']$ yield distinct arrays X. X completely characterizes the probability mass functions $P(S = s, A = k \mid Z = z_j, R = r)$. By Definition A2, $[P_{N_z}(S \mid Z, S^{P_0}), P_{N_z}(S^{P_0} \mid R)^T, P_{N_z}(A \mid S^{P_0})]$ are identifiable. Furthermore, given that $P_{N_z}(S^{P_0} \mid R) P_{N_z}(S^{P_0} \mid R)^+ = \mathbf{I}_{2^{N_z}}$ by the full row rank condition on $P_{N_z}(S^{P_0} \mid R)$, the counterfactual distribution $P_{N_z}(y \mid S^{P_0}, Z = z_j, A = k)$ is given by

$$P(y \mid R, Z = z_j, A = k) P_{N_z}(S^{P_0} \mid R)^+ = P_{N_z}(y \mid S^{P_0}, Z = z_j, A = k) \tag{A9}$$

for all k, j. Again, $P(y \mid R, Z = z_j, A = k)$ completely characterizes the observational distribution for y. By the uniqueness of $P_{N_z}(S^{P_0} \mid R)^+$ and Definition A2, $P_{N_z}(y \mid S^{P_0}, Z = z_j, A = k)$ is identifiable. □

Proof. Proof of Theorem 3

Define the three way array X with dimensions $2N_z \times N_a \times N_r$ and $(i, j, r)^{\text{th}}$ element $P(\tilde{S} = \mathbb{1}(i \leq N_z), A = a \mid Z = z_{i - N_z \mathbb{1}(i > N_z)}, R = r)$. Recall that the definition of matrix $P_{N_z}(\tilde{S} \mid Z, S^{P_0})$ requires that column j be

$$\begin{bmatrix} s_j \\ \mathbf{1}_{N_z} - s_j \end{bmatrix}$$

where the i^{th} element of s_j is denoted as s_{ij} and is defined as:

$$s_{ij} = \text{sn}_S^{\varpi_{N_z}(j-1)_i} (1 - \text{sp}_S)^{1 - \varpi_{N_z}(j-1)_i}$$

Let the matrices $P_{N_z}(S^{P_0} \mid R)^T, P_{N_z}(A \mid S^{P_0})$ be defined as in Appendix A. Then

$$P(\tilde{S} = \mathbb{1}(i \leq N_z), A = k \mid Z = z_{i - N_z \mathbb{1}(i > N_z)}, R = r) = \sum_{j=1}^{2^{N_z}} P_{N_z}(\tilde{S} \mid Z, S^{P_0})_{i,j} P_{N_z}(S^{P_0} \mid R)^T_{r,j} P_{N_z}(A \mid S^{P_0})_{k,j}.$$

Given that $\text{sn}_S + \text{sp}_S \neq 1$, as shown in Lemma A4, $k_{P_{N_z}(\tilde{S}|Z,S^{P_0})} = 3$ and $\text{rank}(P_{N_z}(\tilde{S} \mid Z, S^{P_0})) = N_z + 1$. Furthermore, by assumptions stated in Theorem 3, $\text{rank}(P_{N_z}(S^{P_0} \mid R)^T) = 2^{N_z}$ and $P_{N_z}(S^{P_0} \mid R)^T \in \mathbb{R}^{N_r \times 2^{N_z}}$ so by Definition 4, $k_{P_{N_z}(S^{P_0}|R)^T} = 2^{N_z}$. Given that $k_{P_{N_z}(A|S^{P_0})} \geq 2^{N_z} - 1$ as stated in Theorem 3, the conditions in Lemma A12 hold:

$$\min(3, 2^{N_z}) + 2^{N_z} - 1 \geq 2^{N_z} + 2 \tag{A10}$$

$$\min(3, 2^{N_z} - 1) + 2^{N_z} \geq 2^{N_z} + 2 \tag{A11}$$

$$\tag{A12}$$

and

$$\text{rank}(P_{N_z}(S \mid Z, S^{P_0})) + \text{rank}(P_{N_z}(S^{P_0} \mid R)) + \text{rank}(P_{N_z}(A \mid S^{P_0})) \tag{A13}$$

$$\geq N_z + 1 + 2^{N_z} + 2^{N_z} - 1 \tag{A14}$$

$$= N_z + 2^{N_z+1} \tag{A15}$$

by the fact that $\text{rank}(P_{N_z}(A \mid S^{P_0})) \geq k_{(P_{N_z}(A|S^{P_0})}$. Also

$$N_z + 2^{N_z+1} - 2(2^{N_z} - 1) = N_z - 1 \tag{A16}$$

$$\geq \begin{cases} \min(N_z - 2, \text{rank}(P_{N_z}(A \mid S^{P_0})) - k_{(P_{N_z}(A|S^{P_0})}) \\ \min(N_z - 2, 0) \end{cases}. \tag{A17}$$

Given that $P_{N_z}(A \mid S^{P_0})$ has columns that sum to 1, and $P_{N_z}(S^{P_0} \mid R)^T$ has rows that sum to 1, we can apply Lemma A12 to the 3-way array X. Applying Lemma A12 yields that the triple-product decomposition $[P_{N_z}(\tilde{S} \mid Z, S^{P_0}), P_{N_z}(A \mid S^{P_0}), P_{N_z}(S^{P_0} \mid R)^T]$ is unique up to a common column permutation. However, Theorem 3 states the assumption that sn_S, sp_S lie in a common half-interval. By Lemma A5, the only permutation matrix consistent with the column domain of $P_{N_z}(\tilde{S} \mid Z, S^{P_0})$ is the identity matrix. We conclude that the 3-way decomposition of X, $[P_{N_z}(\tilde{S} \mid Z, S^{P_0}), P_{N_z}(A \mid S^{P_0}), P_{N_z}(S^{P_0} \mid R)^T]$, is unique. It follows that two different decompositions $[P_{N_z}(\tilde{S} \mid Z, S^{P_0}), P_{N_z}(A \mid S^{P_0}), P_{N_z}(S^{P_0} \mid R)^T]$ and $[P_{N_z}(\tilde{S} \mid Z, S^{P_0})', P_{N_z}(A \mid S^{P_0})', (P_{N_z}(S^{P_0} \mid R)^T)']$ yield different Xs. By the fact that X is a complete characterization of the data distribution $P(\tilde{S} = s, A = a \mid Z = z_j, R = r)$ and Definition A2 the parameter set $[P_{N_z}(\tilde{S} \mid Z, S^{P_0}), P_{N_z}(A \mid S^{P_0}), P_{N_z}(S^{P_0} \mid R)^T]$ is strictly identifiable.

Define the matrix $P(\tilde{Y} \mid Z, R, A = k)$ with dimensions $N_z \times N_r$ with elements $P(\tilde{Y} = y \mid R = r, Z = z, A = k)$

$$P(\tilde{Y} \mid Z, R, A = k)_{i,r} = P(\tilde{Y} = 1 \mid Z = z_i, R = r, A = k).$$

Let the matrix $P_{N_z}(\tilde{Y} \mid Z, S^{P_0}, A = k)$ be in $\mathbb{R}^{N_z \times 2^{N_z}}$ for all $k \in \{1, \ldots, N_a\}$ with elements

$$P_{N_z}(\tilde{Y} \mid Z, S^{P_0}, A = k)_{i,j} = \varpi_{N_z}(j-1)_i r_Y P(Y = 1 | Z = z_i, S^{P_0} = \varpi_{N_z}(j-1), A = k) + (1 - \text{sp}_Y) \tag{A18}$$

where $r_Y = \text{sp}_Y + \text{sn}_Y - 1$. Then $P(\tilde{Y} = 1 \mid Z = z_i, A = k, R = r) = \sum_{j=1}^{2^{N_z}} P_{N_z}(\tilde{Y} \mid Z, S^{P_0}, A = a)_{i,j} P_{N_z}(S^{P_0} \mid R)_{j,r}$ which can be represented as matrix multiplication:

$$P(\tilde{Y} \mid Z, R, A = a) = P_{N_z}(\tilde{Y} \mid Z, S^{P_0}, A = a) P_{N_z}(S^{P_0} \mid R) \tag{A19}$$

Given our assumption that $P_{N_z}(S^{P_0} \mid R)$ is full row rank, $P_{N_z}(S^{P_0} \mid R) P_{N_z}(S^{P_0} \mid R)^+ = \mathbf{I}_{2^{N_z}}$ and

$$P(\tilde{Y} \mid Z, R, A = a) P_{N_z}(S^{P_0} \mid R)^+ = P_{N_z}(\tilde{Y} \mid Z, A = a, S^{P_0}) \tag{A20}$$

It then follows the definition of $P_{N_z}(\tilde{Y} \mid Z, S^{P_0}, A = a)$ in Equation (A18) that sp_Y is identifiable, as are the parameters $r_Y P(Y = 1 | Z = z_j, S^{P_0} = \varpi_{N_z}(j-1), A = k)$ for all $z_j, j \in \{1, \ldots, 2^{N_z}\}$ and k.

Let any allowable post-infection outcome vaccine efficacy estimand, necessarily where $u_j\, u_l = 1$, be defined as

$$\text{VE}^u_{I,jl}(k) = 1 - \frac{P(Y(z_j) = 1 \mid S^{P_0} = u, A = k)}{P(Y(z_l) = 1 \mid S^{P_0} = u, A = k)}.$$

By Assumptions 1 to 2 $P(Y = 1 | Z = z, S^{P_0} = u, A = k) = P(Y(z) = 1 \mid S^{P_0} = u, A = k)$ for all $z \in \{z_1, \ldots, z_{N_z}\}$. Note that $\text{sp}_Y = 1 - P_{N_z}(\tilde{Y} \mid Z, S^{P_0}, A = k)_{1,1}$ by our definition of $P_{N_z}(\tilde{Y} \mid Z, S^{P_0}, A = k)$ in Equation (A18). Then

$$P(Y = 1 | Z = z, S^{P_0} = u, A = k) = \frac{P_{N_z}(\tilde{Y} \mid Z, S^{P_0}, A = k)_{z,j} - P_{N_z}(\tilde{Y} \mid Z, S^{P_0}, A = k)_{1,1}}{r_Y}$$

where $j = \varpi^{-1}_{N_z}(u) + 1$, so $\text{VE}^u_{I,jl}(k)$ is identifiable. □

Proof. Proof of Corollary 3
By the conditions set forth in Corollary 3 we have that

$$P(\tilde{S} = \mathbb{1}\left(i \leq N_z\right), \tilde{A} = k \mid Z = z_{i-N_z \mathbb{1}(i>N_z)}, R = r) = \sum_{j=1}^{2^{N_z}} P_{N_z}(\tilde{S} \mid Z, S^{P_0})_{i,j} P_{N_z}(S^{P_0} \mid R)^T_{r,j} P_{N_z}(\tilde{A} \mid S^{P_0})_{k,j}.$$

This decomposition holds because of our nondifferential misclassification assumption, namely $\tilde{A} \perp\!\!\!\perp S^{P_0}, \tilde{S}, R, Z \mid A$, which allows for the following complete characterization of $\tilde{A} \mid S^{P_0}$:

$$P(\tilde{A} = k \mid S^{P_0} = u) = \sum_{\ell=1}^{N_z} P(\tilde{A} = k \mid A = \ell) P(A = \ell \mid S^{P_0} = u).$$

Recall that sn_S, sp_S lie in the same half interval of $[0, 1]$, so by the same logic as Proof 11, the distributions $P(\tilde{S} = 1 \mid Z = z, S^{P_0} = u), P(\tilde{A} = k \mid S^{P_0} = u), P(S^{P_0} = u \mid R = r)$ are identifiable. Define the matrix $P(\tilde{Y} \mid Z, R, \tilde{A} = k)$ with dimensions $N_z \times N_r$ with elements $P(\tilde{Y} = y \mid R = r, Z = z, \tilde{A} = k)$

$$P(\tilde{Y} \mid Z, R, \tilde{A} = k)_{i,r} = P(\tilde{Y} = 1 \mid Z = z_i, R = r, \tilde{A} = k).$$

Let the matrix $P_{N_z}(\tilde{Y} \mid Z, S^{P_0}, \tilde{A} = k)$ be defined in the same way as Equation (A18). Then $P(\tilde{Y} = 1 \mid Z = z_i, \tilde{A} = k, R = r) = \sum_{j=1}^{2^{N_z}} P_{N_z}(\tilde{Y} \mid Z, S^{P_0}, \tilde{A} = a)_{i,j} P_{N_z}(S^{P_0} \mid R)_{j,r}$ which can be represented as matrix multiplication:

$$P(\tilde{Y} \mid Z, R, \tilde{A} = k) = P_{N_z}(\tilde{Y} \mid Z, S^{P_0}, \tilde{A} = a) P_{N_z}(S^{P_0} \mid R) \tag{A21}$$

Given our assumption that $P_{N_z}(S^{P_0} \mid R)$ is full row rank,$P_{N_z}(S^{P_0} \mid R) P_{N_z}(S^{P_0} \mid R)^+ = \mathbf{I}_{2^{N_z}}$ and

$$P(\tilde{Y} \mid Z, R, \tilde{A} = k) P_{N_z}(S^{P_0} \mid R)^+ = P_{N_z}(\tilde{Y} \mid Z, \tilde{A} = k, S^{P_0}) \tag{A22}$$

It then follows the definition of $P_{N_z}(\tilde{Y} \mid Z, S^{P_0}, \tilde{A} = k)$ that sp_Y is identifiable, as are the parameters $r_Y P(Y = 1 | Z = z_j, S^{P_0} = \varpi_{N_z}(j-1), \tilde{A} = k)$ for all $j \in \{1, \ldots, 2^{N_z}\}$ and $k \in \{1, \ldots, N_a\}$.

Let any allowable post-infection outcome vaccine efficacy estimand, necessarily where $u_j u_l = 1$, be defined as

$$\text{VE}^u_{I,jl} = 1 - \frac{P(Y(z_j) = 1 \mid S^{P_0} = u)}{P(Y(z_l) = 1 \mid S^{P_0} = u)}.$$

By Assumptions 1 to 2 $P(Y = 1 | Z = z, S^{P_0} = u, \tilde{A} = k) = P(Y(z) = 1 \mid S^{P_0} = u, \tilde{A} = k)$ for all $z \in \{z_1, \ldots, z_{N_z}\}$. Note that $\text{sp}_Y = 1 - P_{N_z}(\tilde{Y} \mid Z, S^{P_0}, \tilde{A} = k)_{1,1}$ by our definition of $P_{N_z}(\tilde{Y} \mid Z, S^{P_0}, \tilde{A} = k)$ in Equation (A18). Then

$$P(Y = 1 | Z = z, S^{P_0} = u, \tilde{A} = k) = \frac{P_{N_z}(\tilde{Y} \mid Z, S^{P_0}, \tilde{A} = k)_{z,j} - P_{N_z}(\tilde{Y} \mid Z, S^{P_0}, \tilde{A} = k)_{1,1}}{r_Y}$$

where $j = \varpi^{-1}_{N_z}(u) + 1$. Then

$$\text{VE}^u_{I,jl} = 1 - \frac{\sum_k \left(P_{N_z}(\tilde{Y} \mid Z, S^{P_0}, \tilde{A} = k)_{z,j} - P_{N_z}(\tilde{Y} \mid Z, S^{P_0}, \tilde{A} = k)_{1,1}\right) P(\tilde{A} = k \mid S^{P_0} = u)}{\sum_k \left(P_{N_z}(\tilde{Y} \mid Z, S^{P_0}, \tilde{A} = k)_{z,l} - P_{N_z}(\tilde{Y} \mid Z, S^{P_0}, \tilde{A} = k)_{1,1}\right) P(\tilde{A} = k \mid S^{P_0} = u)}.$$

E. Proof of VE bounds

Proof of Theorem 2

Proof. Let the three groups be denoted r_1, r_2, r_3. Then $\theta^r_{(0,1)} = p_{1+0r} - \theta^r_{(1,1)}$ and $\theta^r_{(1,0)} = p_{1+1} - \theta^r_{(1,1)}$. The following system of equations can be defined for all r:

$$\begin{aligned} p_{110r} &= p_{1+0r}\beta_2^{(0,1)} + \theta^r_{(1,1)}(\beta_2^{(1,1)} - \beta_2^{(0,1)}) \\ p_{111r} &= p_{1+1r}\beta_1^{(1,0)} + \theta^r_{(1,1)}(\beta_1^{(1,1)} - \beta_1^{(1,0)}) \end{aligned} \qquad r \in \{r_1, r_2, r_3\}. \tag{A1}$$

These equations reduce to

$$p_{111r} = p_{1+1r}\beta_1^{(1,0)} + \frac{\beta_1^{(1,1)} - \beta_1^{(1,0)}}{\beta_2^{(1,1)} - \beta_2^{(0,1)}}(p_{110r} - p_{1+0r}\beta_2^{(0,1)}) \qquad r \in \{r_1, r_2, r_3\}. \tag{A2}$$

Rewriting eq. (A2) as a linear system and making the reparameterization, $\rho = \frac{\beta_1^{(1,1)} - \beta_1^{(1,0)}}{\beta_2^{(1,1)} - \beta_2^{(0,1)}}$, $(\beta_2^{(0,1)})' = -\rho\beta_2^{(0,1)}$:

$$\begin{bmatrix} p_{111r_1} \\ p_{111r_2} \\ p_{111r_3} \end{bmatrix} = \begin{bmatrix} p_{1+1r_1} & p_{110r_1} & p_{1+0r_1} \\ p_{1+1r_2} & p_{110r_2} & p_{1+0r_2} \\ p_{1+1r_3} & p_{110r_3} & p_{1+0r_3} \end{bmatrix} \begin{bmatrix} \beta_1^{(1,0)} \\ \rho \\ (\beta_2^{(0,1)})' \end{bmatrix}. \tag{A3}$$

Let the matrix $\mathbf{P}$ be:

$$\mathbf{P} = \begin{bmatrix} p_{1+1r_1} & p_{110r_1} & p_{1+0r_1} \\ p_{1+1r_2} & p_{110r_2} & p_{1+0r_2} \\ p_{1+1r_3} & p_{110r_3} & p_{1+0r_3} \end{bmatrix}.$$

In order for the system of equations in eq. (A3) to have a solution we need that

$$\det \mathbf{P}^T \neq 0.$$

This is equivalent to eq. (9) of theorem 2. Then $\beta_2^{(0,1)} = -(\beta_2^{(0,1)})'/\rho$ as long as $\rho \neq 0$. This is equivalent to eq. (10) of theorem 2 which coincides with the dot product of

$$\begin{bmatrix} p_{111r_1} \\ p_{111r_2} \\ p_{111r_3} \end{bmatrix}$$

with the second row of $\mathbf{P}^{-1}$. These equations can be solved for $\beta_2^{(0,1)}, \beta_1^{(1,0)}, \frac{\beta_1^{(1,1)} - \beta_1^{(1,0)}}{\beta_2^{(1,1)} - \beta_2^{(0,1)}}$ given group probability conditions set forth in Theorem 2. The solutions are:

$$\begin{aligned} \beta_1^{(1,0)} &= \frac{p_{111r_1}(p_{100r_3}p_{110r_2} - p_{100r_2}p_{110r_3}) + p_{111r_2}(p_{100r_1}p_{110r_3} - p_{100r_3}p_{110r_1}) + p_{111r_3}(p_{100r_2}p_{110r_1} - p_{100r_1}p_{110r_2})}{p_{1+1r_1}(p_{110r_2}p_{100r_3} - p_{110r_3}p_{100r_2}) + p_{1+1r_2}(p_{100r_1}p_{110r_3} - p_{100r_3}p_{110r_1}) + p_{1+1r_3}(p_{100r_2}p_{110r_1} - p_{100r_1}p_{110r_2})} \\ \frac{\beta_1^{(1,1)} - \beta_1^{(1,0)}}{\beta_2^{(1,1)} - \beta_2^{(0,1)}} &= \frac{p_{1+1r_1}(p_{1+0r_2}p_{111r_3} - p_{1+0r_3}p_{111r_2}) + p_{1+1r_2}(p_{1+0r_3}p_{111r_1} - p_{1+0r_1}p_{111r_3}) + p_{1+1r_3}(p_{1+0r_1}p_{111r_2} - p_{1+0r_2}p_{111r_1})}{-(p_{1+1r_1}(p_{110r_2}p_{100r_3} - p_{110r_3}p_{100r_2}) + p_{1+1r_2}(p_{100r_1}p_{110r_3} - p_{100r_3}p_{110r_1}) + p_{1+1r_3}(p_{100r_2}p_{110r_1} - p_{100r_1}p_{110r_2}))} \\ \beta_2^{(0,1)} &= \frac{p_{111r_1}(p_{101r_3}p_{110r_2} - p_{101r_2}p_{110r_3}) + p_{111r_2}(p_{101r_1}p_{110r_3} - p_{101r_3}p_{110r_1}) + p_{111r_3}(p_{101r_2}p_{110r_1} - p_{101r_1}p_{110r_2})}{p_{111r_1}(p_{1+0r_3}p_{1+1r_2} - p_{1+0r_2}p_{1+1r_3}) + p_{111r_2}(p_{1+0r_1}p_{1+1r_3} - p_{1+0r_3}p_{1+1r_1}) + p_{111r_3}(p_{1+0r_2}p_{1+1r_1} - p_{1+0r_1}p_{1+1r_2})} \end{aligned} \tag{A4}$$

The causal estimand can be represented as

$$\mathrm{VE}_I = 1 - \frac{p_{111r} - p_{1+1r}\beta_1^{(1,0)} + \theta^r_{(1,1)}\beta_1^{(1,0)}}{p_{110r} - p_{1+0r}\beta_2^{(0,1)} + \theta^r_{(1,1)}\beta_2^{(0,1)}} \tag{A5}$$

for all $r \in \{r_1, r_2, r_3\}$, which depends on the unidentified parameters $\theta^r_{(1,1)}$. Bounds on $\theta^r_{(1,1)}$ can be derived via the fact that $\theta^r_{(0,1)}, \theta^r_{(0,1)}, \theta^r_{(1,1)}$, and $\phi_{01}^{(1,1)}, \phi_{01}^{(1,1)}, \phi_{11}^{(1,1)}$ each lie in a probability simplex:

$$\mathbb{1}\left(\theta^r_{(0,1)} \geq 0\right)\mathbb{1}\left(\theta^r_{(1,0)} \geq 0\right)\mathbb{1}\left(\theta^r_{(1,1)} \geq 0\right) \times \mathbb{1}\left(1 \geq \theta^r_{(0,1)} + \theta^r_{(1,0)} + \theta^r_{(1,1)}\right). \tag{A6}$$

This constraint yields the bounds for $\theta^r_{(1,1)}$:

$$\mathbb{1}\left(\min(p_{1+0r}, p_{1+1r}) \geq \theta^r_{(1,1)} \geq \max(p_{1+0r} + p_{1+1r} - 1, 0)\right). \tag{A7}$$

The bounds on $\theta^r_{(1,1)}$ are further modified by taking into account the parameter constraints on $\phi^{(1,1)}_{01}, \phi^{(1,1)}_{10}, \phi^{(1,1)}_{11}$.

$$\mathbb{1}\left(\frac{(p_{110r} - \beta^{(0,1)}_2 p_{1+0r}) - (p_{111r} - \beta^{(1,0)}_1 p_{1+1r})}{\theta^r_{(1,1)}} + \phi^{(1,1)}_{10} + \beta^{(0,1)}_2 - \beta^{(1,0)}_1 \geq 0\right) \mathbb{1}\left(\phi^{(1,1)}_{10} \geq 0\right) \tag{A8}$$

$$\times\, \mathbb{1}\left(\frac{p_{111r} - \beta^{(1,0)}_1 p_{1+1r}}{\theta^r_{(1,1)}} - (\phi^{(1,1)}_{10} - \beta^{(1,0)}_1) \geq 0\right) \mathbb{1}\left(1 \geq \frac{p_{110r} - \beta^{(0,1)}_2 p_{1+0r}}{\theta^r_{(1,1)}} + \beta^{(0,1)}_2 + \phi^{(1,1)}_{10}\right). \tag{A9}$$

The combined bounds on $\theta^r_{(1,1)}$ are

$$\mathbb{1}\Bigg(\min(p_{1+0r}, p_{1+1r}) \geq \theta^r_{(1,1)} \geq \tag{A10}$$

$$\max\left(p_{1+0r} + p_{1+1r} - 1, \frac{p_{110r} - \beta^{(0,1)}_2 p_{1+0r}}{1 - \beta^{(0,1)}_2}, \frac{p_{110r} - \beta^{(0,1)}_2 p_{1+0r}}{-\beta^{(0,1)}_2}, \frac{p_{111} - \beta^{(1,0)}_1 p_{1+1}}{-\beta^{(1,0)}_1}, 0\right)\Bigg). \tag{A11}$$

Let ${}_u\theta^r_{(1,1)} = \min(p_{1+0r}, p_{1+1r})$ and

$${}_l\theta^r_{(1,1)} = \max\left(p_{1+0r} + p_{1+1r} - 1, \frac{p_{110r} - \beta^{(0,1)}_2 p_{1+0r}}{1 - \beta^{(0,1)}_2}, \frac{p_{110r} - \beta^{(0,1)}_2 p_{1+0r}}{-\beta^{(0,1)}_2}, \frac{p_{111} - \beta^{(1,0)}_1 p_{1+1}}{-\beta^{(1,0)}_1}, 0\right).$$

Finally, let VE_{Ir} be defined as function of θ:

$$\mathrm{VE}_{Ir}(\theta) = \frac{p_{110r} - p_{111r} - p_{1+0r}\beta^{(0,1)}_2 + p_{1+1r}\beta^{(1,0)}_1 + \theta(\beta^{(0,1)}_2 - \beta^{(1,0)}_1)}{p_{110r} - p_{1+0r}\beta^{(0,1)}_2 + \theta\beta^{(0,1)}_2}. \tag{A12}$$

For each r, let

$$\mathrm{sign}\left(\frac{\partial \mathrm{VE}_{Ir}}{\partial\theta}\right) = \mathrm{sign}(\beta^{(1,0)}_1\beta^{(0,1)}_2(p_{1+0} - p_{1+1}) - \beta^{(1,0)}_1 p_{110} + \beta^{(0,1)}_2 p_{111})$$

Then the bounds on VE_I are for each r:

$$({}_l\mathrm{VE}_{Ir}, {}_u\mathrm{VE}_{Ir}) = \begin{cases} \left(\mathrm{VE}_{Ir}\left({}_l\theta^r_{(1,1)}\right), \mathrm{VE}_{Ir}\left({}_u\theta^r_{(1,1)}\right)\right) & \mathrm{sign}\left(\frac{\partial \mathrm{VE}_{Ir}}{\partial\theta}\right) = 1 \\ \left(\mathrm{VE}_{Ir}\left({}_u\theta^r_{(1,1)}\right), \mathrm{VE}_{Ir}\left({}_l\theta^r_{(1,1)}\right)\right) & \mathrm{sign}\left(\frac{\partial \mathrm{VE}_{Ir}}{\partial\theta}\right) = -1 \end{cases}$$

Then VE_I is an element of the intersection over r:

$$\mathrm{VE}_I \in \bigcap_{r=1}^{N_r} ({}_l\mathrm{VE}_{Ir}, {}_u\mathrm{VE}_{Ir}).$$

F. Kruskal rank properties

In the section that follows, we use properties and several theorems and lemmas that are proven in Kruskal (1977). Where appropriate we will indicate on which pages the proofs of the theorems and lemmas can be found.

Definition A1 (Array triple product). *Let the array triple product with resulting array $X \in \mathbb{R}^{I\times J\times K}$ be defined between matrices $A \in \mathbb{R}^{I\times R}$, $B \in \mathbb{R}^{J\times R}$, $C \in \mathbb{R}^{K\times R}$. The operation is represented as $X = [A, B, C]$. As a result, the $(i, j, k)^{\text{th}}$ element of X, X_{ijk}, is defined the sum of*

three-way-products of elements a_{ir}, b_{jr}, c_{kr}, i.e.:

$$X_{ijk} = \sum_{r=1}^{R} a_{ir} b_{jr} c_{kr}.$$

LEMMA A10 (RANK LEMMA). *Let*

$$H_{AB}(n) = \min_{\text{card}(A')=n} \{\text{rank}(A') + \text{rank}(B')\} - n$$

for an integer n where A' is an n-column subset of the matrix A and B' is the same column-index subset of a matrix B. For any diagonal matrix $D \in \mathbb{R}^{n \times n}$ with rank δ,

$$\text{rank}(ADB^T) \geq H_{AB}(\delta).$$

See proof on p. 121 in Kruskal (1977).

THEOREM A1 (KRUSKAL TRIPLE PRODUCT DECOMPOSITION UNIQUENESS). *Let matrices A, B, C be defined as in Definition A1, with respective ranks r_A, r_B, r_C, and let array X also be defined as in Definition A1. Suppose that $k_A \leq r_A, k_B \leq r_B$, and $k_C \leq r_C$. Then if*

$$r_A + r_B + r_C - (2R + 2) \geq \begin{cases} \min(r_A - k_A, r_B - k_B) \\ \min(r_A - k_A, r_C - k_C) \end{cases},$$

$\min(k_A, k_B) + r_C \geq R + 2$, *and* $\min(k_A, k_C) + r_B \geq R + 2$ *the decomposition $X = [A, B, C]$ is unique up to column permutation matrix P and column scaling Λ, M, N such that $\Lambda M N$ is the identity matrix. In other words, X can be represented as the triple product of any three matrices $[\tilde{A}, \tilde{B}, \tilde{C}]$ such that $[\tilde{A} = A P \Lambda, \tilde{B} = B P M, \tilde{C} = C P N]$. See proof in Kruskal (1977) on page 126.*

LEMMA A11 (UNIQUENESS WITH FIXED MATRIX AND SUM CONDITION). *Suppose A is fixed and C has columns that sum to one, or $\mathbf{1}_{1 \times K} C = \mathbf{1}_{1 \times R}$. If the rank conditions in Theorem A1 on A, B, C hold then $[A, B, C]$ is the unique triple product decomposition of array X.*

Proof. Suppose that $X = [A, B, C]$ and that $[A, \bar{B}, \bar{C}]$ is another decomposition of X, where $\bar{C}$ satisfies the column-sum constraint. Let $r_{\bar{B}}, r_{\bar{C}}$ be the ranks of $\bar{B}$ and $\bar{C}$ respectively. Definition A1 implies that $A\text{diag}(xC)B^T = A\text{diag}(x\bar{C})\bar{B}^T$ for all $x \in \mathbb{R}^{1 \times I}$. If for any $y \in \mathbb{R}^{1 \times K}$ such that $y\bar{C} = 0 \implies yC = 0$ then $\text{col}(C) \subset \text{col}(\bar{C})$, $\text{null}(C) \supset \text{null}(\bar{C})$, and $r_C \leq r_{\bar{C}}$. If $y\bar{C} = 0$ then

$$A\text{diag}(y\bar{C})\bar{B}^T = 0 \implies A\text{diag}(yC)B^T = 0$$

Recall the definition of $H_{AB}(n)$ from Lemma A10. Kruskal (1977) shows that the condition on the ranks and Kruskal ranks above imply the following inequalities (proof omitted):

$$k_A \geq \max(R - r_B + 2, R - r_C + 2), \tag{A1}$$
$$k_B \geq R - r_C + 2, \tag{A2}$$
$$k_C \geq R - r_B + 2, \tag{A3}$$
$$H_{AB}(n) \geq R - r_C + 2 \text{ if } n \geq R - r_C + 2 \tag{A4}$$
$$H_{AC}(n) \geq R - r_B + 2 \text{ if } n \geq R - r_B + 2 \tag{A5}$$
$$H_{BC}(n) \geq 1 \text{ if } n \geq 1 \tag{A6}$$

The inequality eq. (A4) implies that when $H_{AB}(n) < R - r_C + 2$ then $n < R - r_C + 2$. When $n < R - r_C + 2$, the inequalities eqs. (A1) to (A3) and the definition of $H_{AB}(n)$ imply that $H_{AB}(n) = n$. Then

$$\begin{aligned} 0 &= \text{rank}(A\text{diag}(yC)B^T) \\ &\geq H_{AB}(\text{rank}(\text{diag}(yC))) \geq 0, \end{aligned}$$

where the second to last inequality comes from Lemma A10 and the last inequality comes from the definition of $H_{AB}(n)$. This implies $yC = 0$. Let the function $w(y)$ for a generic vector y return

the number of nonzero entries in the vector y. Let v be any vector such that $w(v\bar{C}) \leq R - \bar{K}_0 + 1$. Then we'll show that $w(vC) \leq w(v\bar{C})$.

$$R - r_C + 1 \geq R - \bar{K}_0 + 1 \geq w(v\bar{C}) = \text{rank}(\text{diag}(v\bar{C}) \tag{A7}$$

$$\geq \text{rank}(A\text{diag}(y\bar{C})\bar{B}^T) = \text{rank}(A\text{diag}(yC)B^T) \tag{A8}$$

$$\geq H_{AB}(\text{rank}(\text{diag}(vC)) = H_{AB}(w(vC)). \tag{A9}$$

The final line implies that $H_{AB}(w(vC)) = w(vC)$, which shows that $w(v\bar{C}) \geq w(vC)$ when $R - \bar{K}_0 + 1 \geq w(v\bar{C})$.

Given this condition, Kruskal's permutation lemma (proved on page 134 of Kruskal (1977)) shows that for any matrices C and $\bar{C}$ that satisfy the inequality, $\bar{C} = CP_CN$ where P_C is a permutation matrix and N is a diagonal nonsingular scaling matrix. If we have the stronger condition that every two columns of C are linearly independent then P_C and N are unique. Our matrices satisfy these conditions, so we have that $\bar{C} = CP_CN$, and a similar argument can be used to show $\bar{B} = BP_BM$

Given that we also have the condition that $\mathbf{1}_{1\times K}C = \mathbf{1}_{1\times R}$ and $\mathbf{1}_{1\times K}\bar{C} = \mathbf{1}_{1\times R}$, then this implies that $\bar{C} = CP_C$ because $\mathbf{1}_{1\times K}\bar{C} = \mathbf{1}_{1\times K}CP_CN = \mathbf{1}_{1\times R}N$. The equality only holds if $N = \mathbf{I}_{R\times R}$.

We now have $\bar{C} = CP_C$ and $\bar{B} = BP_BM$. We can apply Kruskal's permutation matrix proof from pages 129-130 in Kruskal (1977) to show that $P_C = P_B = P$. The following two identities hold for any diagonal scaling matrices M, N, any permutation matrix P, and any vector v:

$$M\text{diag}(v)N = \text{diag}(vMN) \tag{A10}$$

$$P\text{diag}(v)P^T = \text{diag}(vP^T). \tag{A11}$$

Given our condition that $X = [A, B, C] = [A, \bar{B}, \bar{C}]$, Definition 4 implies that for all vectors $v \in \mathbb{R}^{1\times J}$:

$$B\text{diag}(vA)C^T = \bar{B}\text{diag}(vA)\bar{C}^T \tag{A12}$$

$$= BPM\text{diag}(vA)P^TC^T \tag{A13}$$

$$= B\text{diag}(vAMP^T)C^T. \tag{A14}$$

The last line follows from applying Equations (A10) to (A11). The equality $B\text{diag}(vA)C^T = B\text{diag}(vAMP^T)C^T$ implies

$$B\text{diag}(v(A - AMP^T))C^T = 0 \tag{A15}$$

for all v. Furthermore,

$$0 = \text{rank}(B\text{diag}(v(A - AMP^T))C^T) \tag{A16}$$

$$\geq H_{BC}(\text{rank}(\text{diag}(v(A - AMP^T))) \geq 0. \tag{A17}$$

The last line follows from Lemma A10. Then using the implication from eq. (A6) that if $H_{BC}(n) < 1 \implies n = 0$, $\text{rank}(\text{diag}(v(A - AMP^T)) = 0$. In other words, $v(A - AMP^T) = 0$ for all v, from which we conclude

$$A = AMP^T.$$

This equality can hold only if M is the identity and if P^T is the identity, because M is a diagonal nonsingular matrix and P^T is a permutation matrix. □

Lemma A12 (Uniqueness with column and row sum conditions). *Suppose B has rows that sum to 1 and C has columns that sum to 1, or $B\mathbf{1}_{R\times 1} = \mathbf{1}_{J\times 1}$, and $\mathbf{1}_{1\times K}C = \mathbf{1}_{1\times R}$. If the rank conditions in Theorem A1 on A, B, C also hold then $[A, B, C]$ is the unique triple product decomposition of array X up to a common column permutation.*

Proof. Suppose that $X = [A, B, C]$ and that $[\bar{A}, \bar{B}, \bar{C}]$ is another decomposition of X, where $\bar{B}, \bar{C}$ satisfy the respective row- and column-sum constraints.

Following the logic set out in the proof of lemma A11, we can show that $\bar{B} = BP_BM$ and $\bar{C} = CP_CN$.

Given that we also have the condition that $\mathbf{1}_{1\times K}C = \mathbf{1}_{1\times R}$ and $\mathbf{1}_{1\times K}\bar{C} = \mathbf{1}_{1\times R}$, then this implies that $\bar{C} = CP_C$ because $\mathbf{1}_{1\times K}\bar{C} = \mathbf{1}_{1\times K}CP_CN = \mathbf{1}_{1\times R}N$ which only equals $\mathbf{1}_{1\times R}$ if $N = \mathbf{I}_{R\times R}$.

Furthermore, if $r_B = R$, the equation $B\nu = \mathbf{1}_{J\times 1}$ has a unique solution in $\nu \in \mathbb{R}^{R\times 1}$, namely $\nu = \mathbf{1}_{R\times 1}$. This implies that M is the identity matrix, as the condition $\bar{B}\mathbf{1}_{R\times 1} = \mathbf{1}_{J\times 1}$ results in:

$$\mathbf{1}_{J\times 1} = \bar{B}\mathbf{1}_{R\times 1} \tag{A18}$$
$$= BP_BM\mathbf{1}_{R\times 1} \tag{A19}$$
$$\implies P_BM\mathbf{1}_{R\times 1} = \mathbf{1}_{R\times 1}. \tag{A20}$$

Given that M is a nonsingular diagonal matrix and P_B is a permutation matrix, M must be the identity to solve the equation $P_BM\mathbf{1}_{R\times 1} = \mathbf{1}_{R\times 1}$.

We now have $\bar{C} = CP_C$ and $\bar{B} = BP_B$. We can use Kruskal's permutation matrix proof (omitted for brevity) to show that $P_C = P_B = P$. Given the identities in Equations (A10) to (A11). and given our condition that $X = [A, B, C] = [\bar{A}, \bar{B}, \bar{C}]$, for all vectors $v \in \mathbb{R}^{1\times J}$:

$$B\text{diag}(vA)C^T = \bar{B}\text{diag}(v\bar{A})\bar{C}^T \tag{A21}$$
$$= BP\text{diag}(v\bar{A})P^TC^T \tag{A22}$$
$$= B\text{diag}(v\bar{A}P^T)C^T. \tag{A23}$$

The equality $B\text{diag}(vA)C^T = B\text{diag}(v\bar{A}P^T)C^T$ implies

$$B\text{diag}(v(A - \bar{A}P^T))C^T = 0 \tag{A24}$$

for all v. Furthermore,

$$0 = \text{rank}(B\text{diag}(v(A - \bar{A}P^T))C^T) \tag{A25}$$
$$\geq H_{BC}(\text{rank}(\text{diag}(v(A - \bar{A}P^T))) \geq 0. \tag{A26}$$

The last line follows from Lemma A10. Then using the implication from eq. (A6) that if $H_{BC}(n) < 1 \implies n = 0$, $\text{rank}(\text{diag}(v(A - \bar{A}P^T)) = 0$ or $v(A - \bar{A}P^T) = 0$ for all v. This further implies that

$$A = \bar{A}P^T$$

or

$$\bar{A} = AP.$$

□

G. Supporting lemmas and definitions from other work

Definition A2 (Parameter identifiability Rothenberg (1971)). *A parameter $\theta \in \Theta$ is identifiable if there does not exist a distinct parameter value $\theta' \in \Theta$ for which the density $f(y \mid \theta) = f(y \mid \theta')$ for all observations y.*

Lemma A13 (Block rank lemmas Tian (2004)). *Let $A \in \mathbb{R}^{m\times n}, B \in \mathbb{R}^{m\times k}, C \in \mathbb{R}^{l\times n}$.*

$$\text{rank}\left(\begin{bmatrix} A & B \\ C & 0 \end{bmatrix}\right) = \text{rank}(B) + \text{rank}(C) + \text{rank}((I - BB^+)A(I - C^+C))$$

If $\text{range}(B) \subseteq \text{range}(A)$ *and* $\text{range}(C^T) \subseteq \text{range}(A^T)$

$$\text{rank}\left(\begin{bmatrix} A & B \\ C & D \end{bmatrix}\right) = \text{rank}(A) + \text{rank}(D - CA^+B)$$

LEMMA A14 (BLOCK RANK LEMMA EXTENSION). *Let* $A \in \mathbb{R}^{m\times n}, B \in \mathbb{R}^{m\times k}, C \in \mathbb{R}^{l\times n}$. *If* $\text{range}(C^T) \subseteq \text{range}(A^T)$

$$\text{rank}\left(\begin{bmatrix} A & A \\ C & D \end{bmatrix}\right) = \text{rank}(A) + \text{rank}(D - C)$$

Proof. Given that $\text{range}(A) \subseteq \text{range}(A)$, we can apply the second block rank lemma from Lemma A13 with $B = A$.

$$\text{rank}\left(\begin{bmatrix} A & A \\ C & D \end{bmatrix}\right) = \text{rank}(A) + \text{rank}(D - CA^+A).$$

By supposition, $\text{range}(C^T) \subseteq \text{range}(A^T)$ and A^+A is the projection matrix onto the column space of A^T. Then $CA^+A = C$, and the statement follows. □

H. DETAILS BEHIND NUMERICAL EXAMPLES

We have three simulation scenarios where we vary the sample size to determine the power: a two-arm trial to determine vaccine efficacy against severe symptoms, a three-arm trial to determine relative vaccine efficacy against severe symptoms, and a two-arm trial to determine vaccine efficacy against transmission. All trials are designed such that the assumptions of Theorem 3 are satisfied, so the three-arm trial includes 8 study sites, and a categorical covariate with 7 levels, and both two-arm trials include 4 study sites, and a categorical covariate with 3 levels. Within each scenario, we allow for the categorical covariate, A, to be measured perfectly or with error. In addition, we assume a 3-level, pretreatment categorical covariate has been measured for each participant. We simulate from the parametric model defined in Section 2.3, which requires that we specify μ_u^r, or the log-odds of belonging to stratum u relative to base stratum u_0 for each study site r. Let the ordered collection of log-odds of being in stratum u relative to stratum $u_{2^{N_z}}$ for the reference covariate level $x = 1$ be $\mu^r = \left(\mu^r_{u_1}, \mu^r_{u_2}, \ldots, \mu^r_{u_{2^{N_z}-1}}, 0\right)$.

Let softmax be the function from $v \in \mathbb{R}^L$ to the $L+1$-dimensional probability simplex, defined elementwise for the i^{th} element as:

$$\text{softmax}(v)_i = \frac{e^{v_i}}{\sum_{l=1}^{L} e^{v_l^r}}$$

and let softmax^{-1} be the inverse function from $\theta \in$ the $L+1$-dimensional simplex to $\mathbb{R}^L$, where the i^{th} element, $i < L+1$ is defined as

$$\text{softmax}(\theta)_i^{-1} = \log(\theta_i) - \log(\theta_{L+1})$$

Let $\theta_u^{r,x} = P(S^{P_0} = u \mid R = r, X = x)$, and let $\theta^{r,x}$ be the ordered vector $(\theta^{r,x}_{u_1}, \theta^{r,x}_{u_1}, \ldots, \theta^{r,x}_{u_{2^{N_z}}})$. For the 2-arm trials, the population principal strata proportions are as follows:

$$\theta^{r,1} \overset{\text{iid}}{\sim} \text{Dirichlet}((91, 5, 0.5, 3.5))\forall r$$

while for the 3-arm trials, the proportions are

$$\theta^{r,1} \overset{\text{iid}}{\sim} \text{Dirichlet}((91, 5, 0.1, 0.1, 0.1, 0.1, 0.1, 3.5))\forall r$$

These parameter settings roughly equate to a cumulative true incidence of 0.05. Recall from Section 2.3 that $\eta^x \in \mathbb{R}^{2^{N_z}}$, so η_u^x is the change in log-odds of belonging to principal stratum u vs. u_0 relative to $x = 1$. We set $\eta^x_{2^{N_z}} = 0$ for identifiability. Then let $\mu_u^r = \text{softmax}^{-1}(\theta^{r,1})$ and

$$\theta^{r,x} = \text{softmax}\left(\mu_u^r + \eta^x\right),$$

where for all $x > 1$

$$\eta_i^x \overset{\text{iid}}{\sim} \text{Normal}(0,1),\ i < 2^{N_z},\ \eta_{2^{N_z}}^x = 0.$$

Let the N_a-vector $a^{u,x}$ be defined elementwise as $a_k^{u,x}$ where $a_k^{u,x} = P(A = k \mid S^{P_0} = u, X = x)$.

$$a^{u,1} \overset{\text{iid}}{\sim} \text{Dirichlet}(2\mathbf{1}_{N_a}) \forall u,$$

and $\nu^u = \text{softmax}^{-1}(a^{u,1})$. Then recall that $\gamma^x \in \mathbb{R}^{N_a}$ such that γ_k^x is the change in log-odds of $A = k$ relative to $A = k_0$, and that $\gamma_{N_a}^x = 0$ for identifiability. Then

$$a^{u,x} = \text{softmax}\left(\nu^u + \gamma^x\right),$$

and for all $x > 1$

$$\gamma_i^x \overset{\text{iid}}{\sim} \text{Normal}(0,1),\ i < N_a,\ \gamma_{N_a}^x = 0.$$

Finally, recall that

$$\log \frac{P(Y(z_j) = 1 \mid S^{P_0} = u, A = k, X = x)}{P(Y(z_j) = 0 \mid S^{P_0} = u, A = k, X = x)} = \alpha_j^u + \delta_{j,k}^u + \omega_j^x,$$

where ω_j^x is the change in log-odds of $Y(z_j) = 1$, all else being equal, compared to $x = 1$. In all of our simulations, $\omega_j^x = (x-1)\log(1.1)$ for all j. For the 2-arm trial example, we let $\alpha_1^{(1,1)} = \log(0.3/0.7), \alpha_2^{(1,1)} = \log(0.3/0.7) + \log(0.4)$, and $\delta_{1,k}^{(1,1)} = (k-1)\log(0.925), \delta_{2,k}^{(1,1)} = (k-1)\log(0.825)$. Further, we let $\alpha_1^{(1,0)} = \log(0.15/0.85)$, $\alpha_2^{(0,1)} = \log(0.2/0.8)$, and $\delta_{1,k}^{(1,0)} = (k-1)\log(0.925)$, and $\delta_{2,k}^{(0,1)} = 0$

For the 3-arm trial example, we let $\alpha_1^{(1,1,1)} = \log(0.3/0.7), \alpha_2^{(1,1,1)} = \log(0.3/0.7), \alpha_3^{(1,1,1)} = \log(0.3/0.7) + \log(0.4)$, and $\delta_{j,k}^{(1,1,1)} = (k-1)\log(0.925)$ for $j = 1,2,3$. Further, we let $\alpha_1^{(1,0,1)} = \log(0.2/0.8)$, $\alpha_3^{(1,0,1)} = \log(0.1/0.9), \alpha_1^{(1,1,0)} = \log(0.3/0.7)$, $\alpha_2^{(1,1,0)} = \log(0.15/0.85), \alpha_2^{(0,1,1)} = \log(0.25/0.75)$, $\alpha_3^{(0,1,1)} = \log(0.08/0.92)$, $\alpha_3^{(0,0,1)} = \log(0.25/0.75)$, $\alpha_2^{(0,1,0)} = \log(0.25/0.75)$, $\alpha_1^{(1,0,0)} = \log(0.1/0.9)$ and $\delta_{j,k}^u = 0$ for all k, $u \in \{(1,0,0),(0,1,0),(1,1,0),(0,0,1),(1,0,1),(0,1,1)\}$, and all allowable j

In the 2- and 3-arm trial examples that pertain to inferring vaccine efficacy against severe symptoms, we set $\text{sn}_S = 0.8, \text{sp}_S = 0.99$ which reflects the sensitivity and specificity of a typical PCR collected via nasopharyngeal swab (Kissler et al., 2021), and $\text{sn}_Y = 0.99, \text{sp}_S = 0.9$ to reflect the fact that most severe illness caused by the pathogen of interest will be reported, but that there are many severe illness episodes that are reported that may be caused by other pathogens. These lead to a true rate of severe illness of 0.01 but a rate of reported severe illness of 0.1. For comparison Monto et al. (2009) symptom reporting data shows that 10% of participants reported at least one severe symptom, but the cumulative incidence was 0.07.

For the transmission study, we use the same settings for sn_S, sp_S and set $\text{sn}_Y = \text{sn}_S, \text{sp}_Y = \text{sp}_S$. In order to generate $\tilde{A} \mid A$ for each participant, we use the following error model for the three arm trial:

$$P(\tilde{A} = a \mid A = a) = 0.5,\ P(\tilde{A} = a+1 \mid A = a) = P(\tilde{A} = a-1 \mid A = a) = 0.25$$

for $a \in \{2, \ldots, 6\}$. When $a = 1$:

$$P(\tilde{A} = a \mid A = a) = 0.5,\ P(\tilde{A} = a+1 \mid A = a) = 0.5,$$

and when $a = 7$

$$P(\tilde{A} = a \mid A = a) = 0.5,\ P(\tilde{A} = a-1 \mid A = a) = 0.5, a = 7.$$

For the two-arm trials, we generate $\tilde{A} \mid A$ from the following probability model when $a = 2$:

$$P(\tilde{A} = a \mid A = a) = 0.75,\ P(\tilde{A} = a+1 \mid A = a) = P(\tilde{A} = a-1 \mid A = a) = 0.125.$$

When $a = 1$

$$P(\tilde{A} = a \mid A = a) = 0.75,\ P(\tilde{A} = a+1 \mid A = a) = 0.25,$$

and when $a = 3$

$$P(\tilde{A} = a \mid A = a) = 0.75,\ P(\tilde{A} = a-1 \mid A = a) = 0.25.$$

These distributions were chosen to reflect the fact that detailed pre-season antibody titer measurements are typically available for participants in influenza vaccination trials. Further discretizing the titer measurements reduces the misclassification probabilities in our model. Let the collection of these conditional probabilities be $p^a_{N_a}$

For each hypothetical participant in a study site $R = r$ in our study we draw data in the following manner

$$\begin{aligned}
Z_i &\overset{\text{iid}}{\sim} \text{Categorical}(\frac{1}{N_z}\mathbf{1}_{N_z}) \\
X_i &\overset{\text{iid}}{\sim} \text{Categorical}(\frac{1}{3}\mathbf{1}_3) \\
S_i^{P_0} \mid R = r, X = x &\overset{\text{iid}}{\sim} \text{Categorical}(\theta^{r,x}) \\
A_i \mid S^{P_0} = u, X = x &\overset{\text{iid}}{\sim} \text{Categorical}(a^{u,x}) \\
Y_i \mid S^{P_0} = u, A = k, X = x, Z = j &\overset{\text{iid}}{\sim} \text{Bernoulli}(\text{inv_logit}(\alpha^u_j + \delta^u_{j,k} + \omega_x)) \\
\tilde{Y}_i \mid Y = y &\overset{\text{iid}}{\sim} \text{Bernoulli}(y\text{sn}_Y + (1-y)(1-\text{sp}_Y)) \\
\tilde{S}_i \mid S^{P_0} = u, Z = j &\overset{\text{iid}}{\sim} \text{Bernoulli}(u_j\text{sn}_S + (1-u_j)(1-\text{sp}_S)) \\
\tilde{A}_i \mid A = a &\overset{\text{iid}}{\sim} \text{Categorical}(p^a_{N_a})
\end{aligned} \tag{A1}$$

and we do this for all sites $R \in \{1, \ldots, N_r\}$.

We fit the model defined in Equation (11), with the following priors:

$$\begin{aligned}
\text{sn}_S &\sim \text{Beta}(0.5, 1, 4, 2) \\
\text{sp}_S &\sim \text{Beta}(0.5, 1, 10, 2) \\
\text{sn}_Y &\sim \text{Beta}(0.5, 1, 5, 2) \\
\text{sp}_Y &\sim \text{Beta}(0.5, 1, 4, 2) \\
\omega^x_j &\sim \text{Normal}(0, 1), 2 \le x \le 3 \\
\nu^u_k &\sim \text{Normal}(0, 1.7^2), \forall u \in \{u_1, u_2, u_{2^{N_z}}\}, 1 \le k < N_a \\
\nu^u_k &\sim \text{Normal}(0, 0.5^2), \forall u \notin \{u_1, u_2, u_{2^{N_z}}\}, 1 \le k < N_a \\
\gamma^x_k &\sim \text{Normal}(0, 0.5^2), 2 \le x \le 3,\ 1 \le k < N_a \\
\mu^r_u &\sim \text{MultiNormal}((1, 1, \mathbf{0}_{2^{N_z}-3}), \mathbf{I}_{2^{N_z}-1}) \\
\eta^x_u &\sim \text{Normal}(0, 0.5^2) \\
\alpha^u_{j,k} &\sim \text{Normal}(0, 1.7^2)
\end{aligned} \tag{A2}$$

where we have reparameterized our $\alpha^u_j + \delta^u_{j,k}$ to $\alpha^u_{j,k}$ and $\text{Beta}(0.5, 1, 4, 2)$ is the shifted, scaled Beta distribution, where the first two arguments define the support of the distribution, and the second two parameters are shape parameters. For example, for sn_S this corresponds to $\chi \sim \text{Beta}(4, 2)$ and $\text{sn}_S = (1 - 0.5)\chi + 0.5$.

We use Stan for inference (Team, 2021) using the `cmdstanr` package (Gabry & Češnovar, 2022) in `R` (R Core Team, 2022). All models were run for 6,000 warmup and 6,000 post-warmup iterations; all $\hat{R}$ statistics (Gelman et al., 2013) were below 1.01, as recommended by Vehtari et al. (2020). The bulk and tail effective sample sizes were greater than 9% of samples for all models.

In order to ensure that our decision rule did not lead to high Type 1 error rates, we simulated data under the null hypothesis that vaccine efficacy against post-infection outcomes was 0. The results presented in Table 3 show that the Type 1 error rates are less that 0.05 for each scenario.

Table 3. *Type* 1 *error rates for vaccine efficacy against severe illness designs*

Trial	A meas.	4,000	20,000	40,000	80,000	120,000
3-arm	A	NA	NA	0.00	0.02	0.00
	$\tilde{A}$	NA	NA	0.00	0.03	0.01
2-arm	A	0.03	0.01	0.00	0.04	NA
	$\tilde{A}$	0.01	0.02	0.04	0.03	NA

The Type 1 error rates for vaccine efficacy against transmission are presented in Table 4. None of the Type 1 error rates are statistically significantly greater than 0.05.

Table 4. *Type* 1 *error rates for vaccine efficacy against transmission designs*

A	0.03	0.08	0.05	0.04
$\tilde{A}$	0.00	0.09	0.03	0.03

References

Allman, E. S., Matias, C. & Rhodes, J. A. (2009). Identifiability of parameters in latent structure models with many observed variables. *The Annals of Statistics* **37**.

Cheng, J. & Small, D. S. (2006). Bounds on causal effects in three-arm trials with non-compliance. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **68**.

Ding, P., Geng, Z., Yan, W. & Zhou, X.-H. (2011). Identifiability and Estimation of Causal Effects by Principal Stratification With Outcomes Truncated by Death. *Journal of the American Statistical Association* **106**.

Ding, P. & Lu, J. (2017). Principal stratification analysis using principal scores. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **79**.

Frangakis, C. E. & Rubin, D. B. (2002). Principal Stratification in Causal Inference. *Biometrics* **58**.

Gabry, J. & Češnovar, R. (2022). *cmdstanr: R Interface to 'CmdStan'*. Https://mc-stan.org/cmdstanr/, https://discourse.mc-stan.org.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A. & Rubin, D. B. (2013). Bayesian data analysis .

Gilbert, P. B., Bosch, R. J. & Hudgens, M. G. (2003). Sensitivity Analysis for the Assessment of Causal Vaccine Effects on Viral Load in HIV Vaccine Trials. *Biometrics* **59**.

Grilli, L. & Mealli, F. (2008). Nonparametric Bounds on the Causal Effect of University Studies on Job Opportunities Using Principal Stratification. *Journal of Educational and Behavioral Statistics* **33**.

Gustafson, P. (2015). *Bayesian inference for partially identified models: Exploring the limits of limited data*, vol. 140. CRC Press.

Hudgens, M. G. & Halloran, M. E. (2006). Causal Vaccine Effects on Binary Postinfection Outcomes. *Journal of the American Statistical Association* **101**.

Jemiai, Y., Rotnitzky, A., Shepherd, B. E. & Gilbert, P. B. (2007). Semiparametric estimation of treatment effects given base-line covariates on an outcome measured after a post-randomization event occurs: Semiparametric Estimation of Treatment Effects. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **69**.

Jiang, Z. & Ding, P. (2020). Measurement errors in the binary instrumental variable model. *Biometrika* **107**.

Jiang, Z., Ding, P. & Geng, Z. (2016). Principal causal effect identification and surrogate end point evaluation by multiple trials. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **78**.

Kissler, S. M., Fauver, J. R., Mack, C., Olesen, S. W., Tai, C., Shiue, K. Y., Kalinich, C. C., Jednak, S., Ott, I. M., Vogels, C. B. F., Wohlgemuth, J., Weisberger, J., DiFiori, J., Anderson, D. J., Mancell, J., Ho, D. D., Grubaugh, N. D. & Grad, Y. H. (2021). Viral dynamics of acute SARS-CoV-2 infection and applications to diagnostic and public health strategies. *PLOS Biology* **19**.

Kruskal, J. B. (1977). Three-way arrays: Rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics. *Linear Algebra and its Applications* **18**.

Lipsitch, M. & Kahn, R. (2021). Interpreting vaccine efficacy trial results for infection and transmission. *Vaccine* **39**.

Long, D. M. & Hudgens, M. G. (2013). Sharpening Bounds on Principal Effects with Covariates: Principal Effect Bounds. *Biometrics* **69**.

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

Monto, A. S., Ohmit, S. E., Petrie, J. G., Johnson, E., Truscon, R., Teich, E., Rotthoff, J., Boulton, M. & Victor, J. C. (2009). Comparative Efficacy of Inactivated and Live Attenuated Influenza Vaccines. *New England Journal of Medicine* **361**.

Ouyang, J. & Xu, G. (2022). Identifiability of Latent Class Models with Covariates. *Psychometrika* .

Polack, F. P., Thomas, S. J., Kitchin, N., Absalon, J., Gurtman, A., Lockhart, S., Perez, J. L., Pérez Marc, G., Moreira, E. D., Zerbini, C., Bailey, R., Swanson, K. A., Roychoudhury, S., Koury, K., Li, P., Kalina, W. V., Cooper, D., Frenck, R. W., Hammitt, L. L., Türeci, Ö., Nell, H., Schaefer, A., Ünal, S., Tresnan, D. B., Mather, S., Dormitzer, P. R., Şahin, U., Jansen, K. U. & Gruber, W. C. (2020). Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine. *New England Journal of Medicine* **383**.

R Core Team (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rothenberg, T. J. (1971). Identification in Parametric Models. *Econometrica* **39**.

Rubin, D. B. (2006). Causal Inference Through Potential Outcomes and Principal Stratification: Application to Studies with "Censoring" Due to Death. *Statistical Science* **21**.

Shepherd, B. E., Gilbert, P. B., Jemiai, Y. & Rotnitzky, A. (2006). Sensitivity Analyses Comparing Outcomes Only Existing in a Subset Selected Post-Randomization, Conditional on Covariates, with Application to HIV Vaccine Trials. *Biometrics* **62**.

Shepherd, B. E., Gilbert, P. B. & Lumley, T. (2007). Sensitivity Analyses Comparing Time-to-Event Outcomes Existing Only in a Subset Selected Postrandomization. *Journal of the American Statistical Association* **102**.

Shi, X., Miao, W., Nelson, J. C. & Tchetgen Tchetgen, E. J. (2020). Multiply robust causal inference with double-negative control adjustment for categorical unmeasured confounding. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **82**.

Tchetgen Tchetgen, E. J. (2014). Identification and estimation of survivor average causal effects. *Statistics in Medicine* **33**.

Team, S. D. (2021). *Stan Modeling Language Users Guide and Reference Manual, v2.27*.

Tenforde, M. W., Self, W. H., Adams, K., Gaglani, M., Ginde, A. A., McNeal, T., Ghamande, S., Douin, D. J., Talbot, H. K., Casey, J. D., Mohr, N. M., Zepeski, A., Shapiro, N. I., Gibbs, K. W., Files, D. C., Hager, D. N., Shehu, A., Prekker, M. E., Erickson, H. L., Exline, M. C., Gong, M. N., Mohamed, A., Henning, D. J., Steingrub, J. S., Peltan, I. D., Brown, S. M., Martin, E. T., Monto, A. S., Khan, A., Hough, C. L., Busse, L. W., ten Lohuis, C. C., Duggal, A., Wilson, J. G., Gordon, A. J., Qadir, N., Chang, S. Y., Mallow, C., Rivas, C., Babcock, H. M., Kwon, J. H., Halasa, N., Chappell, J. D., Lauring, A. S., Grijalva, C. G., Rice, T. W., Jones, I. D., Stubblefield, W. B., Baughman, A., Womack, K. N., Rhoads, J. P., Lindsell, C. J., Hart, K. W., Zhu, Y., Olson, S. M., Kobayashi, M., Verani, J. R., Patel, M. M. & Influenza and Other Viruses in the Acutely Ill (IVY) Network (2021). Association Between mRNA Vaccination and COVID-19 Hospitalization and Disease Severity. *JAMA* .

Tian, Y. (2004). Rank equalities for block matrices and their moore-penrose inverses. *Houston Journal of Mathematics* **30**, 483–510.

VanderWeele, T. J. & Tchetgen Tchetgen, E. J. (2011). Bounding the Infectiousness Effect in Vaccine Trials. *Epidemiology* **22**.

Vehtari, A., Gelman, A., Simpson, D., Carpenter, B. & Bürkner, P.-C. (2020). Rank-normalization, folding, and localization: An improved rhat for assessing convergence of mcmc. *Bayesian Analysis* .

Wang, W., Xu, Y., Gao, R., Lu, R., Han, K., Wu, G. & Tan, W. (2020). Detection of SARS-CoV-2 in Different Types of Clinical Specimens. *JAMA* .

World Health Organization (2017). Evaluation of Influenza Vaccine Effectiveness: A Guide to the Design and Interpretation of Observational Studies .

Zhang, J. L. & Rubin, D. B. (2003). Estimation of Causal Effects via Principal Stratification When Some Outcomes are Truncated by "Death". *Journal of Educational and Behavioral Statistics* **28**.

Zhang, J. L., Rubin, D. B. & Mealli, F. (2009). Likelihood-Based Analysis of Causal Effects of Job-Training Programs Using Principal Stratification. *Journal of the American Statistical Association* **104**.