

Research Article

Peter B. Gilbert*, Bryan S. Blette, Bryan E. Shepherd, and Michael G. Hudgens

Post-randomization Biomarker Effect Modification Analysis in an HIV Vaccine Clinical Trial

<https://doi.org/10.1515/jci-2019-0022>

Received Aug 05, 2019; accepted Apr 06, 2020

Abstract: While the HVTN 505 trial showed no overall efficacy of the tested vaccine to prevent HIV infection over placebo, markers measuring immune response to vaccination were strongly correlated with infection. This finding generated the hypothesis that some marker-defined vaccinated subgroups were partially protected whereas others had their risk increased. This hypothesis can be assessed using the principal stratification framework (Frangakis and Rubin, 2002) for studying treatment effect modification by an intermediate response variable, using methods in the sub-field of principal surrogate (PS) analysis that studies multiple principal strata. Unfortunately, available methods for PS analysis require an augmented study design not available in HVTN 505, and make untestable structural risk assumptions, motivating a need for more robust PS methods. Fortunately, another sub-field of principal stratification, survivor average causal effect (SACE) analysis (Rubin, 2006) – which studies effects in a single principal stratum – provides many methods not requiring an augmented design and making fewer assumptions. We show how, for a binary intermediate response variable, methods developed for SACE analysis can be adapted to PS analysis, providing new and more robust PS methods. Application to HVTN 505 supports that the vaccine partially protected individuals with vaccine-induced T-cells expressing certain combinations of functions.

Keywords: Principal stratification; Randomized clinical trial; Two-phase sampling

1 Introduction

A vaccine that prevents HIV-1 infection is critically needed for ending the global HIV-1 pandemic. A recent efficacy trial of a candidate HIV-1 vaccine – HVTN 505 – randomized 2,496 HIV-1 negative volunteers in 1:1 allocation to the DNA/rAd5 vaccine regimen or placebo [1] administered at months 0, 1, 3, 6. The primary objective compared the rate of HIV-1 infection from 6.5 to 24 months between the randomized treatment arms, with estimated cumulative incidence 4.62% (3.15%) in the vaccine (placebo) group, and cumulative incidence ratio 1.46 (95% CI 0.82 to 2.63, Wald test $p = 0.20$). Through a 2-phase sampling design, Janes et al. [2] studied HIV-1 Envelope-specific CD8+ T cell responses measured 2 weeks post last vaccination (Month 6.5) as a correlate of HIV-1 infection between 6.5 and 24 months, and found in vaccine recipients a strong inverse correlation between CD8+ T cell polyfunctionality score (PFS) and HIV-1 infection ($p < 0.001$). The surprising strength of the correlate [e.g., estimated cumulative risk 0.160 and 0.034 for vaccinated subgroups with PFS below and above the median response, compared to 0.070 for placebo recipients (Figure 4 of [2]), suggests the possibil-

***Corresponding Author: Peter B. Gilbert:** Department of Biostatistics, University of Washington and Fred Hutchinson Cancer Research Center, Seattle, Washington, 98109, United States of America; Email: pgilbert@fredhutch.org

Bryan S. Blette, Michael G. Hudgens: Department of Biostatistics, University of North Carolina, Chapel Hill, North Carolina, 27599, United States of America

Bryan E. Shepherd: Department of Biostatistics, Vanderbilt University School of Medicine, Nashville, Tennessee, 37232, United States of America

ity that a qualitative interaction occurred, where some marker-defined vaccinated subgroups received partial protection from vaccination whereas others had their risk increased.

Janes et al.'s observation of intermediate risk of the placebo group between that of the biomarker response-defined vaccinated subgroups does not imply causal vaccine vs. placebo effect modification across these subgroups, because post-randomization selection bias could occur (e.g., due to an unmeasured genetic factor predictive of both the PFS and HIV infection). A direct assessment of a causal vaccine effect (eliminating possible selection bias) would compare risk for each vaccinated subgroup defined by biomarker response value to that of the placebo recipient subgroup who would have had the same biomarker response value if assigned vaccination, which essentially repeats the primary analysis of vaccine efficacy (which is valid based on the randomization) across the marker-defined subgroups. This "principal surrogate (PS) analysis" [3] is a sub-field of the general framework of principal stratification established by Frangakis and Rubin [3].

However, available methods for PS analysis make strong structural risk assumptions that do not hold in HVTN 505, and require study design augmentations such as close-out placebo vaccination to measure counterfactual biomarker response [4] that were not utilized in HVTN 505. Given the high threshold of evidence needed to convince scientists of a qualitative interaction, application of these methods would be inadequate – rather PS analysis with methods that rely on weaker assumptions are needed. Fortunately, many nonparametric and semiparametric principal stratification methods are available that do not require a design augmentation, many of which were developed for applications in another sub-field of principal stratification – "survivor average causal effect (SACE) analysis" (e.g., [5, 6]) – which we define as principal stratification methods focused on a single principal stratum (e.g., always survivors). We describe how such methods designed for SACE analysis can be adapted to PS analysis for the special case of a binary intermediate outcome. This adaptation contributes novel PS methods, by (1) applying to clinical trials without design augmentations; (2) relaxing the strong assumption of no individual-level clinical treatment effects before the biomarker is measured; (3) avoiding strong structural placebo conditional risk assumptions; and (4) extending SACE methods to study designs that only measure the intermediate outcome in a subset of participants (e.g., two-phase/case-control studies), as in HVTN 505. The PFS marker studied by Janes et al. [2] is a summary measure of many different cellular expression patterns, which are binary (expressed vs. not expressed), motivating methods for a binary intermediate outcome.

2 Principal Surrogate (PS) Target Parameters and the HVTN 505 Application

A common objective of randomized clinical trials is to assess biomarker response endpoints as principal surrogate endpoints for the clinical endpoint of interest [3]. With Z denoting binary treatment assignment and S^τ denoting a biomarker measured at a fixed time point τ post-randomization and Y denoting the binary clinical endpoint of interest measured after τ , the PS estimand of interest is the "principal effects" or "causal effect predictiveness" (CEP) surface [7]: $CEP(s_1, s_0) = h(risk_1(s_1, s_0), risk_0(s_1, s_0))$, where $risk_z(s_1, s_0) \equiv P(Y(z) = 1 | S^\tau(1) = s_1, S^\tau(0) = s_0, Y^\tau(1) = Y^\tau(0) = 0)$ for $z = 0, 1$. Here $Y^\tau(z)$ is the indicator that the clinical endpoint occurs by time τ , such that the population for inference is free of the endpoint under both treatment assignments when S^τ is measured. In HVTN 505, S^τ measures an immune response to vaccination and $Y = 1$ is subsequent HIV-1 infection, and it would be problematic to include individuals already infected with HIV-1, because their S^τ values would be affected by the natural immune response to HIV-1 infection, making the results uninterpretable. Therefore, $CEP(s_1, s_0)$ contrasts the risk of infection between the vaccine and placebo arms among those who would be infection-free at time τ under either treatment assignment and have biomarker measures s_1 and s_0 under assignment to vaccine and placebo, respectively. The objective of PS analysis is inference about the $CEP(s_1, s_0)$ parameters and contrasts in these parameters, which has been addressed using a variety of methodological approaches including estimated maximum likelihood [4, 7, 8], pseudo-score estimating equations [9], principal score methods [10–12], and Bayesian methods [13–15].

We consider PS methods assuming either “No Early Effect” (**NEE**), “No Early Harm” (**NEH**), or “No Early Benefit” (**NEB**) of treatment on Y by time τ , as defined below as **A4**, **A4'**, and **A4''**. For the PS methods assuming **NEH** or **NEB**, we focus on the special case that $Z = 0$ is a control condition such as placebo, and there is no variability of S^τ in subjects assigned to $Z = 0$, i.e., $P(S^\tau(0) = 0 | Y^\tau(0) = 0) = 1$. This “Constant Biomarker (CB)” case occurs in placebo-controlled preventive vaccine efficacy trials that only enroll subjects not previously infected with the pathogen under study and for which the intermediate response endpoint is a readout from a validated bioassay designed to only detect an immune response *specific* to the pathogen under study [7]; HVTN 505 is a typical example. For methods assuming **NEE**, we consider both the special Case CB and the general case where $S^\tau(0)$ varies and a monotonicity assumption holds (**A5** below). These scenarios are chosen because they commonly occur in applications. Table 1 describes the interpretation of the $CEP(s_1, s_0)$ parameters for HVTN 505. Because Case CB holds, $CEP(1, 0)$ and $CEP(0, 0)$ are of interest. For a given binary immune response biomarker S^τ , $VE(1) = CEP(1, 0)$ is vaccine efficacy for the subgroup with positive response if assigned vaccine and negative response if assigned placebo, and $VE(0) = CEP(0, 0)$ is vaccine efficacy for the subgroup with negative response under both treatment assignments. Our scientific goal is to study whether vaccine efficacy differs between these two subgroups, i.e. $\mu \equiv VE(1) - VE(0) \neq 0$, and we also seek to assess evidence for a qualitative interaction, defined by $VE(1)$ and $VE(0)$ having opposite signs.

Table 1: Notation and Parameters of Interest for the HVTN 505 HIV-1 Vaccine Efficacy Trial

Variable	Description
Z	Randomized treatment assignment to vaccine ($z = 1$) or placebo ($z = 0$)
Y	Indicator of diagnosis of HIV-1 infection after $\tau = 6.5$ months post enrollment through 24 months (outcome of interest)
Y^τ	Indicator of diagnosis of HIV-1 infection by $\tau = 6.5$ months
S^τ	Binary immune response biomarker measured at τ (eligible if $Y^\tau = 0$)
$risk_z(s_1, s_0)$	$= P(Y(z) = 1 S^\tau(1) = s_1, S^\tau(0) = s_0, Y^\tau(1) = Y^\tau(0) = 0)$: Probability of infection diagnosis if assigned $Z=z$ among the subgroup uninfected at τ under either treatment assignment and with biomarker measures s_1 and s_0 under assignment to vaccine and placebo, respectively. HVTN 505 studies binary biomarkers S^τ described in Section 7.
$CEP(s_1, s_0)$	$= h(risk_1(s_1, s_0), risk_0(s_1, s_0))$ with $h(x, y)$ a contrast function satisfying $h(x, y) = 0$ if and only if $x = y$ and $h(x, y) < 0$ for $x < y$.
Case CB	A study where S^τ is constant (“Constant Biomarker”) in $Z = 0$ participants.
$VE(s_1)$	$= CEP(s_1, 0)$ for $s_1 \in \{0, 1\}$ with $h(x, y) = 1 - x/y$. In HVTN 505, Case CB holds ($P(S^\tau = 0) = 1$). $VE(1)$ and $VE(0)$ are vaccine efficacy for the subgroups with biomarker response if assigned vaccine $S^\tau(1) = 1$ and $S^\tau(1) = 0$, respectively.

3 Assumptions and Identifiability of the PS Target Parameters $CEP(s_1, s_0)$

3.1 Additional Notation and Identifiability Assumptions

Let W be a vector of baseline covariates measured in everyone, and R be the indicator of whether the binary endpoint S^τ is measured at τ . To fit the most common applications in clinical trials we assume Y is binary; in the case of a time-to-event outcome, $Y = I(T \leq t)$, where T is the time from randomization until the endpoint

and $t > \tau$ is a fixed time point of interest. Note that S^τ is undefined if $T \leq \tau$, which we denote as $S^\tau = *$. Define $p(s_1, s_0) \equiv P(S^\tau(1) = s_1, S^\tau(0) = s_0 | Y^\tau(1) = Y^\tau(0) = 0)$ for $(s_1, s_0) \in \{(0, 0), (1, 0), (1, 1)\}$.

Throughout we make a baseline set of assumptions made in essentially all previous frequentist inference SACE papers. We assume the $(Z_i, W_i, R_i, Y_i^\tau(1), Y_i^\tau(0), S_i^\tau(1), S_i^\tau(0), Y_i(1), Y_i(0))$, $i = 1, \dots, n$, are iid [with observed data $O_i = (Z_i, W_i, R_i, S_i^\tau, Y_i^\tau, Y_i)$, with S_i^τ only observed if $R_i = 1$] and assume **A1** SUTVA (*Stable Unit Treatment Value Assumption*); **A2** Ignorable Treatment Assignment: Conditional on W , Z is independent of $(Y^\tau(1), Y^\tau(0), S^\tau(1), S^\tau(0), Y(1), Y(0))$; and **A3** No Censoring or Random Censoring: The binary endpoint Y is observed for all participants, or, if $Y = I(T \leq t)$ with T subject to right-censoring C , $C(z)$ is random conditional on W ($T(z) \perp C(z) | W$ for $z = 0, 1$). As introduced above, we develop the methods under each of three assumptions regarding early final endpoint events before the intermediate endpoint S^τ is measured at τ :

A4 No Early Effect (**NEE**): $P(Y^\tau(1) = Y^\tau(0)) = 1$

A4' No Early Harm (**NEH**): $P(Y^\tau(1) \leq Y^\tau(0)) = 1$

A4'' No Early Benefit (**NEB**): $P(Y^\tau(1) \geq Y^\tau(0)) = 1$

We also assume $p(1, 0) > 0$, which holds trivially for any PS application of interest. **A4'** and **A4''** are standard SACE monotonicity assumptions for the analysis of $\{risk_1, risk_0\}$, which weakens the strong **NEE** assumption made in almost all papers on PS methods.

For the non-Case CB scenario where the biomarker $S^\tau(0)$ varies we also reduce the number of non-identified terms via a biomarker monotonicity assumption:

A5 Biomarker Monotonicity: $P(S^\tau(0) \leq S^\tau(1) | Y^\tau(1) = Y^\tau(0) = 0) = 1$.

A5 states that among participants who will not experience the clinical outcome $Y = 1$ by τ regardless of treatment assignment, none have a negative treatment effect on the biomarker.

We develop the methods under four scenarios of assumption sets— **NEE-VB**: **A1–A5** hold and Variable Biomarker (**VB**) $S^\tau(0)$; **NEE-CB**: **A1–A4** hold and Constant Biomarker (**CB**); **NEH-CB**: **A1–A3**, **A4'** hold and Constant Biomarker; **NEB-CB**: **A1–A3**, **A4''** hold and Constant Biomarker.

3.2 Identifiability of $CEP(s_1, s_0)$

There are 16 basic principal strata corresponding to all possible combinations of values for $(S^\tau(1), S^\tau(0), Y^\tau(1), Y^\tau(0))$. Our assumptions, corresponding to the four scenarios stated above, imply that many of these basic principal strata are empty. Our interest lies in parameters defined for four principal strata: three defined by the basic principal strata $S^\tau(1) = s_1, S^\tau(0) = s_0, Y^\tau(1) = Y^\tau(0) = 0$ with $(s_1, s_0) \in (0, 0), (1, 0)$, and $(1, 1)$, and the fourth defined by $Y^\tau(1) = Y^\tau(0) = 0$. Identification of these parameters requires additional assumptions that can be expressed using sensitivity parameters. Once sensitivity parameters are fixed, the parameters of interest become identifiable. Derivations for identification are given in Web Appendix A. For scenarios **NEE-VB**, **NEE-CB**, **NEH-CB**, and **NEB-CB**, the numbers of required sensitivity parameters are 2, 1, 4, and 2.

4 SACE Methods and their Translation to PS Analysis

4.1 SACE Methods

In this section we use generic notation (Z, S, Y) to define the general SACE parameter [16], and in Section 4.2 show how it maps to the notation for the PS problem. A common objective is to assess the effect of a randomized binary treatment Z on an outcome Y , where Y is only defined or observable for participants with

a post-randomization intermediate response variable S equal to a certain level ($S = 1$, say). The SACE is the average causal effect of Z on Y in the subgroup with $S = 1$ under both treatment assignments:

$$\text{SACE} \equiv h(E[Y(1)|S(1) = S(0) = 1], E[Y(0)|S(1) = S(0) = 1]), \quad (1)$$

where $h(x, y)$ is a contrast function meeting the requirements described in Table 1. Here $S(z)$ and $Y(z)$ are potential outcomes of S and Y if a subject is assigned to treatment $Z = z$.

The large SACE methods literature has focused on the additive-difference contrast $h(x, y) = x - y$ and includes techniques for nonparametric bounds [17–21] and techniques for sensitivity analysis that estimate the SACE under a spectrum of selection bias models [22–27]. Several of these methods make the SACE monotonicity assumption [$P(S(0) \leq S(1)) = 1$] and several relax this assumption. We use a general contrast function $h(x, y)$ because vaccine efficacy is typically measured by a multiplicative reduction in risk rather than an additive difference.

4.2 SACE and PS Methods as Two Strains of Principal Stratification

For PS applications, the multiple SACE-defining intermediate outcomes $S = 1$ depend on both Y^τ and S^τ (detailed in Table 2). Because S^τ is typically categorical or continuous, PS applications makes inference across a spectrum of principal stratum subgroups, whereas SACE applications make inference for a single principal stratum. Focusing on a single stratum has facilitated development of nonparametric and semiparametric SACE methods that need only deal with one or a few terms that are not identified from the observed data, whereas an attempt to apply SACE methods for each principal stratum defined by a general categorical S^τ would face a much larger number of non-identified terms that grows with the number of categories. However, there is an important special case for which SACE methods can be practically adapted—when S^τ is binary—for which only a few principal strata are of interest. The motivating HVTN 505 application has binary markers of interest.

Under each of the four assumption sets defined above, we show how to apply SACE methods to the PS problem: inference about $CEP(0, 0)$, $CEP(1, 0)$, and $CEP(1, 1)$. First, we write the overall conditional risks, $\text{risk}_z \equiv P(Y(z) = 1 | Y^\tau(1) = Y^\tau(0) = 0)$ for $z = 0, 1$, as weighted averages of the marker-subgroup conditional risks:

$$\text{risk}_z = p(0, 0)\text{risk}_z(0, 0) + p(1, 0)\text{risk}_z(1, 0) + p(1, 1)\text{risk}_z(1, 1) \quad \text{for } z = 0, 1. \quad (2)$$

Equation (2) uses the fact that $p(0, 1) = 0$ under the assumptions we use.

The risk_z and $\text{risk}_z(s_1, s_0)$ parameters measure risks in marker-defined subsets of the “always early uninfected” principal stratum defined by $\{Y^\tau(1) = Y^\tau(0) = 0\}$. For scenario **NEH-CB** we will also use the following parameters measuring risks in subsets of the “early protected” (EP) principal stratum defined by $\{Y^\tau(1) = 0, Y^\tau(0) = 1\}$: $\text{risk}_z^{EP}(s_1, s_0) \equiv P(Y(z) = 1 | S^\tau(1) = s_1, S^\tau(0) = s_0, Y^\tau(1) = 0, Y^\tau(0) = 1)$ for $z, s_1, s_0 \in \{0, 1\}$. And, for **NEB-CB**, the analogous parameters are needed for the “early harmed” principal stratum.

The observation that motivated this work is that a contrast in risk_1 and risk_0 is a SACE (defined by intermediate event $S = 1 - Y^\tau = 1$), a contrast in $\text{risk}_1(0, 0)$ and $\text{risk}_0(0, 0)$ is a SACE (defined by intermediate event $S = [1 - Y^\tau][1 - S^\tau] = 1$), a contrast in $\text{risk}_1(1, 1)$ and $\text{risk}_0(1, 1)$ is a SACE (defined by intermediate event $S = [1 - Y^\tau]S^\tau = 1$), and, whereas a contrast in $\text{risk}_1(1, 0)$ and $\text{risk}_0(1, 0)$ is not a SACE, these two parameters are identified from (2) and the other parameters. Thus any two existing SACE methods can be employed to estimate the two sets of means/probabilities $\{\text{risk}_1, \text{risk}_0\}$ and $\{\text{risk}_1(0, 0), \text{risk}_0(0, 0)\}$, and another for $\{\text{risk}_1(1, 1), \text{risk}_0(1, 1)\}$ in scenario **NEE-VB**, and then equation (2) is applied to yield estimates of the remaining two probabilities $\{\text{risk}_1(1, 0), \text{risk}_0(1, 0)\}$ via

$$\text{risk}_z(1, 0) = [\text{risk}_z - p(0, 0)\text{risk}_z(0, 0) - p(1, 1)\text{risk}_z(1, 1)] / p(1, 0) \quad (3)$$

(where $\text{risk}_z(1, 1) = 0$ in the Case CB scenarios **NEE-CB**, **NEH-CB**, and **NEB-CB**).

We interpret the SACE parameters in terms of our motivating application. The first is $SACE_{AEU} = h(risk_1, risk_0)$, with the AEU subgroup being “always early uninfected” individuals with $S = 1$ indicating being uninfected with HIV-1 at time τ . The second is $SACE_{AEUB0} = h(risk_1(0, 0), risk_0(0, 0))$, with AEUB0 being “always early uninfected and biomarker (B) value $S^\tau = 0$ ” individuals. The third is $SACE_{AEUB1} = h(risk_1(1, 1), risk_0(1, 1))$, for “always early uninfected and biomarker value $S^\tau = 1$ ” individuals.

Table 2: Use of SACE Methods [With Numbers of Sensitivity Parameters (S.P.s)] for Estimation of $CEP(s_1, s_0) = h(risk_1(s_1, s_0), risk_0(s_1, s_0))$ for $(s_1, s_0) = (0, 0), (1, 0), (1, 1)$

Target Parameter	Assumption Sets (Identifiability Assumptions)			
	NEE-VB (A1–A5)	NEE-CB (A1–A4)	NEH-CB (A1–A3, A4')	NEB-CB (A1–A3, A4'')
$SACE_{AEU} = h(risk_1, risk_0)$	Nonpar. ident.	Nonpar. ident.	Any SACE method w/ mon. A4' ^a	Any SACE method w/ mon. A4'' ^b
SACE with $S \equiv 1 - Y^\tau$	(0 S.P.s)	(0 S.P.s)	(1 S.P.)	(1 S.P.)
$SACE_{AEUB0} = h(risk_1(0, 0), risk_0(0, 0))$	Any SACE method w/ mon. A5 ^{c,d}	Any SACE method w/ mon. A5 ^d	Any SACE method w/o mon. ^e	Any SACE method w/ mon. ^f
SACE with $S \equiv [1 - Y^\tau][1 - S^\tau]$	(1 S.P.)	(1 S.P.)	(3 S.P.s)	(1 S.P.)
$SACE_{AEUB1} = h(risk_1(1, 1), risk_0(1, 1))$	Any SACE method w/ mon. A5 ^g	N/A	N/A	N/A
SACE with $S \equiv [1 - Y^\tau]S^\tau$	(1 S.P.)	N/A	N/A	N/A

^aThe monotonicity assumption for the SACE method is $P(S(0) \leq S(1)) = 1$ with $S \equiv 1 - Y^\tau$.

^bThe monotonicity assumption for the SACE method is $P(S(0) \geq S(1)) = 1$ with $S \equiv 1 - Y^\tau$.

^cThe monotonicity assumption is $P(S(1) \leq S(0)) = 1$ with $S \equiv [1 - Y^\tau][1 - S^\tau]$.

^dBy **A4**, the monotonicity assumption in footnote c simplifies to $P(S^\tau(0) \leq S^\tau(1) | Y^\tau = 0) = 1$, which holds by **A5** in scenario **NEE-VB** and by Case CB in scenario **NEE-CB**.

^eUnder **A4'** and Case CB in scenario **NEH-CB**, the monotonicity assumption expressed in footnote c amounts to no $Z = 1$ participants with a negative biomarker response at τ . This assumption is often difficult to justify and hence we use methods without monotonicity.

^fThe monotonicity assumption expressed in footnote c holds in scenario **NEB-CB** by **A4''**.

^gThe monotonicity assumption is $P(S(0) \leq S(1)) = 1$ with $S \equiv [1 - Y^\tau]S^\tau$. By **A4**, the monotonicity assumption expressed in footnote e simplifies to $P(S^\tau(0) \leq S^\tau(1) | Y^\tau = 0) = 1$, which holds by **A5** in scenario **NEE-VB**.

The CEP parameters are linked to the three SACE parameters and risks through $CEP(0, 0) = SACE_{AEUB0}$, $CEP(1, 1) = SACE_{AEUB1}$, and $CEP(1, 0) = h(risk_1(1, 0), risk_0(1, 0))$ with $risk_z(s_1, s_0)$ as in (3). Thus our approach pieces together SACE estimates and estimates of the means $p(s_1, s_0)$ to yield estimates of the CEP parameters, where the estimators focusing on subgroups defined by S^τ must account for S^τ only measured in a subset, e.g., through inverse probability weighting. For $h(x, y) = x - y$, $h(risk_1(1, 0), risk_0(1, 0)) = [SACE_{AEU} - p(0, 0) SACE_{AEUB0} - p(1, 1) SACE_{AEUB1}] / p(1, 0)$.

4.3 Mapping Existing SACE Methods to Estimate the PS Parameters $CEP(s_1, s_0)$

Table 2 summarizes how SACE methods can be used to estimate the $CEP(s_1, s_0)$ parameters for the four assumption-set scenarios. Under **NEE-VB**, $SACE_{AEU}$ is nonparametrically identified by **A4**. $CEP(0, 0) =$

$SACE_{AEUB0}$ can be estimated using any SACE method making the monotonicity assumption $P(S(1) \leq S(0)) = 1$, with $S = [1 - Y^\tau] [1 - S^\tau] = 1$. Similarly, $CEP(1, 1) = SACE_{AEUB1}$ can be estimated with a SACE method using the same monotonicity assumption but now with $S = [1 - Y^\tau] S^\tau = 1$. The mixing parameters, $p(s_1, s_0)$, are all nonparametrically identified, such that $CEP(1, 0) = h(risk_1(1, 0))$, $risk_0(1, 0)$ can be estimated based on equation (2).

Estimation under **NEE-CB** is identical to that of scenario **NEE-VB** except that $risk_1(1, 1)$ and $risk_0(1, 1)$ vanish (because the principal stratum defined by $[1 - Y^\tau(0)] S^\tau(0) = [1 - Y^\tau(1)] S^\tau(1) = 1$ is empty). For scenario **NEH-CB**, $SACE_{AEU}$ can be estimated using an existing SACE method making the monotonicity assumption $P(S(1) \leq S(0)) = 1$, with intermediate event $S = 1 - Y^\tau = 1$. Then, $CEP(0, 0) = SACE_{AEUB0}$ can be estimated using an arbitrary SACE method that does not assume monotonicity.

5 Illustration of the Approach with IPW Extensions of Particular SACE Methods

For each assumption set scenario, we show how the PS estimation and inference works using an extension of Shepherd, Gilbert, and Dupont's [24] SACE method, which in scenarios **NEE-VB**, **NEE-CB** simplifies to the Gilbert, Bosch, and Hudgens [22] (GBH) SACE method. The extension incorporates inverse probability weighting (IPW) to handle missing S^τ . We focus on semiparametric efficient estimators given the data $(Z, R, Y^\tau, S^\tau, Y)$ not including baseline covariates W , which amount to sample means with or without IPW as needed. Moreover, we focus on the case that Y is binary and observed for all participants; Web Appendix C summarizes how the methods translate to $Y = I(T \leq t)$ with T subject to right-censoring before t . We use general estimating function notation so that users preferring to use more efficient estimators leveraging information in W (e.g., [28]) may substitute alternative estimating functions into the equations.

5.1 General IPW Estimation

The SACE estimators involve estimation of identified terms $E[Y|S = 1, Z = z]$ for subgroups $S = 1$ [with $S = 1 - Y^\tau$, $(1 - Y^\tau)S^\tau$, or $(1 - Y^\tau)(1 - S^\tau)$] and of terms $E[S^\tau|Y^\tau = 0, Z = z]$, where S^τ is measured at time τ and may be missing. Define the probability of observing S^τ as $\pi(O_1) \equiv P(R = 1|O_1)$, where O_1 is data observed in everyone, i.e., (Z, W, Y^τ, Y) . We assume S^τ is missing at random, $\pi(O_1) = P(R = 1|O_1, S^\tau)$, and that $\pi(O_1)$ is bounded away from zero, $\pi(O_1) \geq \varepsilon$ with probability 1 for some fixed $\varepsilon > 0$.

Following standard IPW estimation, we specify a model $\pi(O_1, \psi)$ for $\pi(O_1)$ (e.g., logistic), and estimate the parameter ψ by maximum likelihood, yielding $\hat{\pi}_i = \pi(O_{1i}, \hat{\psi})$. Efficiency and robustness may be improved by calibrating the estimated weights $\hat{\pi}_i$ accounting for W_i (e.g., [29, 30]).

5.2 Dichotomous Outcome SACE Methods Under Scenario NEE-VB

For scenario **NEE-VB**, the first step is to estimate the terms that are nonparametrically identified— $\{risk_1, risk_0\}$, $p(0, 0)$, $p(1, 0)$, and $p(1, 1)$. Each $risk_z$ for $z = 0, 1$ can be estimated by any preferred method for estimating a mean, most simply by solving $\sum_{i=1}^n U_i^{0z}(O_i; risk_z) = 0$ with estimating function $U_i^{0z}(O_i; risk_z) \equiv (1 - Y_i^\tau)I(Z_i = z)(Y_i - risk_z)$.

Given $p(0, 0) = P(S^\tau = 0|Z = 1, Y^\tau = 0)$, if full data were available, then a simple approach would estimate $p(0, 0)$ by solving $\sum_{i=1}^n U_i^{01}(O_i; p(0, 0)) = 0$ with $U_i^{01}(O_i; p(0, 0)) \equiv (1 - Y_i^\tau)Z_i(1 - S_i^\tau - p(0, 0))$, with convention $U_i^{01}(O_i; p(0, 0)) = 0$ if $S_i^\tau = *$. The IPW version of this equation is $\sum_{i=1}^n R_i U_i^{01}(O_i; p(0, 0))/\hat{\pi}_i = 0$. The parameter $p(1, 1)$ may be estimated similarly by solving $\sum_{i=1}^n R_i U_i^{00}(O_i; p(1, 1))/\hat{\pi}_i = 0$ with $U_i^{00}(O_i; p(1, 1)) \equiv (1 - Y_i^\tau)(1 - Z_i)(S_i^\tau - p(1, 1))$, again with convention $U_i^{00}(O_i; p(1, 1)) = 0$ if $S_i^\tau = *$. With $\hat{p}(0, 0)$ and $\hat{p}(1, 1)$, we then set $\hat{p}(1, 0) = 1 - \hat{p}(0, 0) - \hat{p}(1, 1)$.

Lastly, we estimate each pair $\{risk_1(0, 0), risk_0(0, 0)\}$ and $\{risk_1(1, 1), risk_0(1, 1)\}$ with a SACE method that assumes monotonicity. We summarize how the GBH method, extended to incorporate IPW, accomplishes this task.

GBH Semiparametric SACE Method with IPW (Notation as in GBH). Consider the odds ratio selection bias model with user-specified fixed sensitivity parameter β_0 :

$$\mathbf{B.0} \quad \exp(\beta_0) = \frac{P(S(1)=1|S(0)=1, Y(0)=1)/\{1-P(S(1)=1|S(0)=1, Y(0)=1)\}}{P(S(1)=1|S(0)=1, Y(0)=0)/\{1-P(S(1)=1|S(0)=1, Y(0)=0)\}}.$$

Under **A1–A3**, **B.0**, monotonicity $[P(S(1) \leq S(0)) = 1]$, and positivity $[P(S(1) = 0, S(0) = 1) > 0, P(S(1) = 1, S(0) = 1) > 0]$, the two parameters of interest, $P^{11}(z) \equiv P(Y(z) = 1|S(1) = S(0) = 1)$ for $z = 0, 1$, are nonparametrically identified.

By monotonicity $P^{11}(1) = P(Y(1) = 1|S(1) = 1)$, such that $P^{11}(1)$ is estimated by solving $\sum_{i=1}^n R_i U^1(O_i; P(Y(1) = 1|S(1) = 1))/\hat{\pi}_i = 0$ where $U^1(O_i; P(Y(1) = 1|S(1) = 1)) = Z_i S_i (Y_i - P(Y(1) = 1|S(1) = 1))$. Next, $P^{11}(0)$ is estimated by first estimating α_0 as the solution to $\sum_{i=1}^n R_i U^0(O_i; \alpha_0, \beta_0)/\hat{\pi}_i = 0$ where

$$U^0(O_i; \alpha_0, \beta_0) = Z_i \left(S_i - \hat{P}(S(0) = 1) \sum_{y=0}^1 w_0(y; \alpha_0, \beta_0) \hat{P}(Y(0) = y|S(0) = 1) \right), \quad (4)$$

where $\hat{P}(Y(0) = 1|S(0) = 1)$ is obtained in the same way as $\hat{P}(Y(1) = 1|S(1) = 1)$. Then with $\hat{\alpha}_0$ from (4), $\hat{P}^{11}(0) = [\hat{P}(S(0) = 1)/\hat{P}(S(1) = 1)]w_0(1; \hat{\alpha}_0, \beta_0)\hat{P}(Y(0) = 1|S(0) = 1)$. We implement this “IPW GBH SACE Method” verbatim multiple times below, with the meaning of $S = 1$ changing for estimating needed terms in $CEP(s_1, s_0)$.

Assumptions Needed for Valid Inference with the IPW GBH SACE Method. We state additional assumptions, beyond the identifiability assumptions, that are needed for valid inference via the IPW GBH SACE Method. The method requires $P(S(1) = 1) < P(S(0) = 1)$ in order that $\hat{P}^{11}(z)$ for each $z = 0, 1$ has an asymptotic normal distribution and thus to ensure that Wald confidence intervals for $P^{11}(z)$ based on asymptotic or nonparametric bootstrap variance estimates have correct coverage probabilities [23]. Moreover, if $P(S(1) = 1) < P(S(0) = 1)$ but the probabilities are close, then the Wald confidence intervals can have poor coverage. The general SACE assumption $P(S(1) = 1) < P(S(0) = 1)$ translated for ensuring valid inference on $SACE_{AEUB0}$ and $SACE_{AEUB1}$ for **NEE-VB** (Table 2) is **A6** “Early Biomarker Effect”: $P(S^\tau(0) = 1|Y^\tau = 0) < P(S^\tau(1) = 1|Y^\tau = 0)$. **A6** is also the assumption needed for valid inference on $SACE_{AEUB0}$ for **NEE-CB**. For **NEH-CB**, the general SACE assumption translated for valid inference on $SACE_{AEU}$ is **A7** “Early Benefit”: $P(Y^\tau(1) = 1) < P(Y^\tau(0) = 1)$, while for **NEB-CB** it is **A7'** “Early Harm” $P(Y^\tau(1) = 1) > P(Y^\tau(0) = 1)$. Lastly, for **NEB-CB** the general SACE assumption translated for inference on $SACE_{AEUB0}$ is **A8**: $P(S^\tau(1) = 0|Y^\tau(1) = 0) < P(Y^\tau(0) = 0)/P(Y^\tau(1) = 0)$. **A6** almost always holds in applications and one of **A7** or (**A7'**, **A8**) plausibly holds for many real binary PS applications, as discussed in Section 7. Moreover, **A7**, **A7'**, **A8** are testable such that the conditions needed to assure valid inference can be checked.

Implementing the IPW GBH SACE Method for $CEP(s_1, s_0)$. The semiparametric MLEs $\hat{risk}_1(0, 0)$ and $\hat{risk}_0(0, 0)$ are obtained as $\hat{P}^{11}(1)$ and $\hat{P}^{11}(0)$, using $S \equiv [1 - Y^\tau][1 - S^\tau]$ with **B.0** and a fixed β_0 . The semiparametric MLEs $\hat{risk}_1(1, 1)$ and $\hat{risk}_0(1, 1)$ are obtained in the same way with $S \equiv [1 - Y^\tau]S^\tau$, with monotonicity assumption in the reverse direction [i.e., $P(S(1) \geq S(0)) = 1$]. This means that we use the IPW GBH SACE Method reversing the roles of $Z = 1$ and $Z = 0$, leading to a selection bias model defined by:

$$\mathbf{B.1} \quad \exp(\beta_1) = \frac{P(S(0)=1|S(1)=1, Y(1)=1)/\{1-P(S(0)=1|S(1)=1, Y(1)=1)\}}{P(S(0)=1|S(1)=1, Y(1)=0)/\{1-P(S(0)=1|S(1)=1, Y(1)=0)\}}.$$

The estimate $\hat{P}^{11}(0)$ is obtained based on $U^0(O_i; P(Y(0) = 1|S(0) = 1)) = (1 - Z_i)S_i(Y_i - P(Y(0) = 1|S(0) = 1))$ and $\hat{P}^{11}(1)$ is obtained by setting

$$\hat{P}^{11}(1) = [\hat{P}(S(1) = 1)/\hat{P}(S(0) = 1)]w_1(1; \hat{\alpha}_1, \beta_1)\hat{P}(Y(1) = 1|S(1) = 1) \quad (5)$$

after estimating α_1 from $U^1(O_i; \alpha_1, \beta_1) =$

$$(1 - Z_i) \left(S_i - \hat{P}(S(1) = 1) \sum_{y=0}^1 w_1(y; \alpha_1, \beta_1) \hat{P}(Y(1) = y|S(1) = 1) \right). \quad (6)$$

By standard estimating equation theory, the above estimators are consistent and asymptotically normal for given fixed β_0 and β_1 . To obtain Wald confidence intervals for each $CEP(s_1, s_0)$, consistent estimating-function based variance estimators may be used for the estimates $\hat{P}^{11}(z)$ not involving $\hat{\alpha}_0$ or $\hat{\alpha}_1$; e.g., the estimated variance of $\hat{P}^{11}(1)$ for the IPW GBH SACE Method is given by $\sum_{i=1}^n (R_i/\hat{\pi}_i) \left[U^1(O_i; \hat{P}(Y(1) = 1|S(1) = 1)) \right]^2$. Influence-function based variance estimates are similarly obtained for the estimates $\hat{P}^{11}(z)$ involving $\hat{\alpha}_0$ or $\hat{\alpha}_1$, by using a vector estimating function and the delta method. For example, the four components of the estimating function in (6) are for $(\hat{\alpha}_1, \hat{P}(S(1) = 1), \hat{P}(S(0) = 1), \hat{P}(Y(1) = 1|S(1) = 1))^T$, with delta method applied with $g(w, x, y, z) = \frac{x}{y} w_1(1; w, \beta_1)z$. All variance estimation is performed with the R package geex [31].

To perform a sensitivity analysis, one approach specifies a plausible range $[l_k, u_k]$ (or maximum possible) for each sensitivity parameter β_k , $k = 0, 1$. An ignorance interval for $CEP(s_1, s_0)$ may be estimated by the span of values between the minimum and maximum estimates, obtained by setting β_0 and β_1 to the boundary values. Using the method of Imbens and Manski [32] and Vansteelandt et al. [33], a Wald asymptotic $(1-\alpha)\%$ estimated uncertainty interval (EUI) for $CEP(s_1, s_0)$ may be calculated as in formulas (40) and (41) of Richardson et al. [34], using the variance estimates of the minimum and maximum $CEP(s_1, s_0)$ estimates. In particular, let $\hat{CEP}_l(s_1, s_0)$ and $\hat{CEP}_u(s_1, s_0)$ be the estimates of $CEP(s_1, s_0)$ fixing the sensitivity parameters at the values within a pre-specified plausible region $\Gamma = [l_0, u_0] \times [l_1, u_1]$ of the sensitivity parameter(s) that minimize or maximize $\hat{CEP}(s_1, s_0)$, respectively. With $\hat{\sigma}_l^2$ and $\hat{\sigma}_u^2$ consistent estimates of the asymptotic limiting variances of $\hat{CEP}_l(s_1, s_0)$ and $\hat{CEP}_u(s_1, s_0)$, respectively, a $(1-\alpha)\%$ EUI is given by $[\hat{CEP}_l(s_1, s_0) - c_\alpha \hat{\sigma}_l / \sqrt{n}, \hat{CEP}_u(s_1, s_0) + c_\alpha \hat{\sigma}_u / \sqrt{n}]$, where c_α satisfies

$$\Phi \left(c_\alpha + \left(\sqrt{n}(\hat{CEP}_u(s_1, s_0) - \hat{CEP}_l(s_1, s_0)) \right) / \max\{\hat{\sigma}_l, \hat{\sigma}_u\} \right) - \Phi(-c_\alpha) = 1 - \alpha,$$

where $\Phi(\cdot)$ denotes the cdf of a standard normal variate. The same approach can be used to construct Wald confidence intervals and EUIs for the other scenarios and SACE approaches described below. Theoretical justification of these EUIs relies on the assumption that the values $\gamma_l, \gamma_u \in \Gamma$ that correspond to the ignorance interval for $CEP(s_1, s_0)$ are the same for all possible observed data laws (condition (39) from [34]), which holds for **NEE-VB** and **NEE-CB** and may need validation for **NEH-CB** and **NEB-CB** applications.

5.3 Dichotomous Outcome SACE Methods Under Scenario NEE-CB

For scenario **NEE-CB**, $CEP(s_1, 0)$ for $s_1 = 0, 1$ can be estimated exactly as for **NEE-VB**, with one change that $SACE(1, 1)$ vanishes because $p(1, 1) = 0$. In particular, first $\{risk_1, risk_0\}$ and $p(0, 0)$ are estimated as in scenario **NEE-VB**, and then $p(1, 0)$ is estimated as $\hat{p}(1, 0) = 1 - \hat{p}(0, 0)$. Secondly, $\{risk_1(0, 0), risk_0(0, 0)\}$ are estimated as in scenario **NEE-VB**. Lastly, $risk_z(1, 0)$ for each $z = 0, 1$ is estimated via equation (3) plugging in estimates for each term. These steps amount to first estimating $risk_1(0, 0)$ by the solution to $\sum_{i=1}^n R_i Z_i (1 - Y_i^T)(1 - S_i^T)(Y_i - risk_1(0, 0)) / \hat{\pi}_i = 0$, with convention that the summand is zero if $S_i^T = *$. Then $risk_0(0, 0)$ and $risk_0(1, 0)$ are estimated by the solutions to the equations **B.0** and $\hat{risk}_0 - risk_0(0, 0)\hat{p}(0, 0) - risk_0(1, 0)\hat{p}(1, 0) = 0$; our code for the simulation study and example are implemented in this manner.

5.4 Dichotomous Outcome SACE Methods Under Scenario NEH-CB

We implement an IPW extension of the Shepherd, Gilbert, and Dupont [24] SACE method that relaxes monotonicity by using the sensitivity parameter β_0 in **B.0** plus three additional sensitivity parameters:

$$\mathbf{B.2} \quad \exp(\beta_2) = \frac{risk_1(0, 0) / \{1 - risk_1(0, 0)\}}{risk_1(0, *) / \{1 - risk_1(0, *)\}}$$

$$\mathbf{B.3} \quad \exp(\beta_3) = \frac{risk_1(1, 0) / \{1 - risk_1(1, 0)\}}{risk_1(1, *) / \{1 - risk_1(1, *)\}}$$

$$\mathbf{B.4} \quad \exp(\beta_4) = \frac{p(1, 0) / \{1 - p(1, 0)\}}{P(S^T(1)=1|0, 1) / \{1 - P(S^T(1)=1|0, 1)\}},$$

where $risk_1(s_1, *) \equiv P(Y(1) = 1 | S^T(1) = s_1, S^T(0) = *, Y^T(1) = 0, Y^T(0) = 1)$ for $S_1 = 0, 1$ and $P(S^T(1) = 1 | 0, 1) \equiv P(S^T(1) = 1 | Y^T(1) = 0, Y^T(0) = 1)$. The estimation steps are similar to those taken for scenario **NEE-CB**: First, estimate $p(1, 0)$ and $P(S^T(1) = 1 | Y^T(1) = 0, Y^T(0) = 1)$ as the solutions to the two equations **B.4** and

$$\begin{aligned} & \hat{P}(S^T(1) = 1 | Y^T(1) = 0) - p(1, 0) \hat{P}(Y^T(0) = 0 | Y^T(1) = 0) \\ & - P(S^T(1) = 1 | Y^T(1) = 0, Y^T(0) = 1) \{1 - \hat{P}(Y^T(0) = 0 | Y^T(1) = 0)\} = 0. \end{aligned}$$

Then $risk_0(0, 0)$ and $risk_0(1, 0)$ are estimated as in scenario **NEE-CB**. Finally, $risk_1(s_1, 0)$ and $risk_1(s_1, *)$ are estimated as the solutions to $\hat{P}(Y(1) = 1 | Y^T(1) = 0, S^T(1) = s_1) - risk_1(s_1, 0) \hat{P}(Y^T(0) = 0, S^T(0) = 0 | Y^T(1) = 0, S^T(1) = s_1) - risk_1(s_1, *) \{1 - \hat{P}(Y^T(0) = 0, S^T(0) = 0 | Y^T(1) = 0, S^T(1) = s_1)\} = 0$ and either **B.2** for $s_1 = 0$ or **B.3** for $s_1 = 1$.

5.5 Dichotomous Outcome SACE Methods Under Scenario NEB-CB

A SACE method for this scenario uses the sensitivity parameter β_0 in **B.0** and one additional sensitivity parameter defined as **B.5**:

$$\mathbf{B.5} \quad \exp(\beta_5) = \frac{risk_0/(1-risk_0)}{P(Y(0)=1 | Y^T(1)=1, Y^T(0)=0)/(1-P(Y(0)=1 | Y^T(1)=1, Y^T(0)=0))}.$$

The parameters $CEP(1, 0)$ and $CEP(0, 0)$ can be estimated exactly as in **NEE-CB**, except that $risk_0$ is not identifiable and thus its estimation leverages the additional assumption **B.5**. We estimate $risk_0$ and $P(Y(0) = 1 | Y^T(1) = 1, Y^T(0) = 0)$ as the solutions to **B.5** and $P(Y(0) = 1 | Y^T(0) = 0) = P(Y^T(1) = 0 | Y^T(0) = 0) risk_0 + P(Y^T(1) = 1 | Y^T(0) = 0) P(Y(0) = 1 | Y^T(1) = 1, Y^T(0) = 0)$. Then, estimation of $risk_1$, $p(0, 0)$, $risk_1(1, 0)$, $risk_1(0, 0)$, $risk_0(1, 0)$, and $risk_0(0, 0)$ is identical to that described for **NEE-CB**.

5.6 Effect Modification Analysis: Inference on Contrasts in $CEP(s_1, s_0)$

For making inference on contrasts in the $CEP(s_1, s_0)$, such as $\mu \equiv CEP(1, 0) - CEP(0, 0)$, our set-up constrains μ to values narrower than the maximum possible range -2 to 2 . For example, for scenario **NEE-CB**, setting $\beta_0 = 0$ implies $risk_0(1, 0) = risk_0(0, 0)$, which leaves each of $CEP(1, 0)$ and $CEP(0, 0)$ free to vary over the maximum possible range as for any SACE method but constrains μ to -1 to 1 . Thus making inference on contrasts of $CEP(s_1, s_0)$ does not achieve just-nonparametric identifiability as does inference on the individual $CEP(s_1, s_0)$ parameters. This should be borne in mind when inference is made on CEP contrasts as well as on the individual parameters.

6 Simulation Study

We first simulated data sets under the assumptions of **NEE-CB**. For each of n independent individuals, we generated potential outcomes and then observed outcomes. First, $(Y^T(1), Y^T(0))$ was set to $(0, 0)$ or $(1, 1)$ with probabilities 0.8 and 0.2 , such that **A4 (NEE)** holds. If $Y^T(1) = 1$, then $Y(0)$ and $Y(1)$ were set to 1 . If $Y^T(1) = 0$, then $(S^T(1), S^T(0))$ was set to $(0, 0)$ or $(1, 0)$ with probabilities 0.4 and 0.6 , such that Case CB holds.

To evaluate size and power of a test of $H_0 : CEP(1, 0) = CEP(0, 0)$ versus $H_1 : CEP(1, 0) \neq CEP(0, 0)$, data were simulated under 13 different values for $CEP(1, 0) - CEP(0, 0)$: $-0.6, -0.5, \dots, 0.6$. Specifically, if $Y^T(1) = 0$ and $S^T(1) = S^T(0) = 0$, then $Y(1)$ was generated as Bernoulli with mean a ($Bern(a)$) and $Y(0)$ as $Bern(0.5)$. If $Y^T(1) = 0$ and $S^T(1) = 1, S^T(0) = 0$, then $Y(1)$ was $Bern(b)$ and $Y(0)$ was $Bern(0.5)$. Thus $CEP(1, 0) - CEP(0, 0) = b - a$ for contrast function $h(x, y) = x - y$. The values of a ranged from 0.7 to 0.1 by decrements of 0.05 and b increased from 0.1 to 0.7 by increments of 0.05 . Under this parameterization $risk_0(0, 0) = 0.5 = risk_0(1, 0)$, implying $\beta_0 = 0$.

To generate the observed data, Z was drawn from $\text{Bern}(0.5)$, and the vector of observable random variables $(Z, Y^T, S^T, Y) = (Z, Y^T(Z), S^T(Z), Y(Z))$ was determined. Simulations were conducted with and without case-cohort sampling of S^T . For the latter, membership in the random subcohort was determined by R drawn from $\text{Bern}(\nu)$. For individuals with $Y^T = 0$, S^T was observed for subcohort members (i.e., $R = 1$) and cases (i.e., $Y = 1$).

Data sets were generated for all 156 combinations of: $n \in \{200, 400, 800, 1600\}$; $\nu \in \{0.1, 0.25, 1\}$; and $\text{CEP}(1, 0) - \text{CEP}(0, 0) \in \{-0.6, -0.5, \dots, 0.6\}$, except the 13 scenarios with $n = 200$ and $\nu = 0.1$ were not studied because only 8 vaccine recipient uninfected controls are expected to have S^T measured. All results are based on 2000 simulated data sets. Analyses used $[l_0, u_0] \in \{[0, 0], [-1, 1], [-2.5, 2.5]\}$. The true values for $\text{CEP}(1, 0)$ and $\text{CEP}(0, 0)$ were selected to include zero effect modification and gradients in effect modification magnitudes in both directions, with values $(a, b) \in \{(0.1, 0.7), (0.15, 0.65), (0.2, 0.6), (0.25, 0.55), (0.55, 0.25), (0.6, 0.2), (0.65, 0.15), (0.7, 0.1)\}$ expressing qualitative interactions. For each simulated data

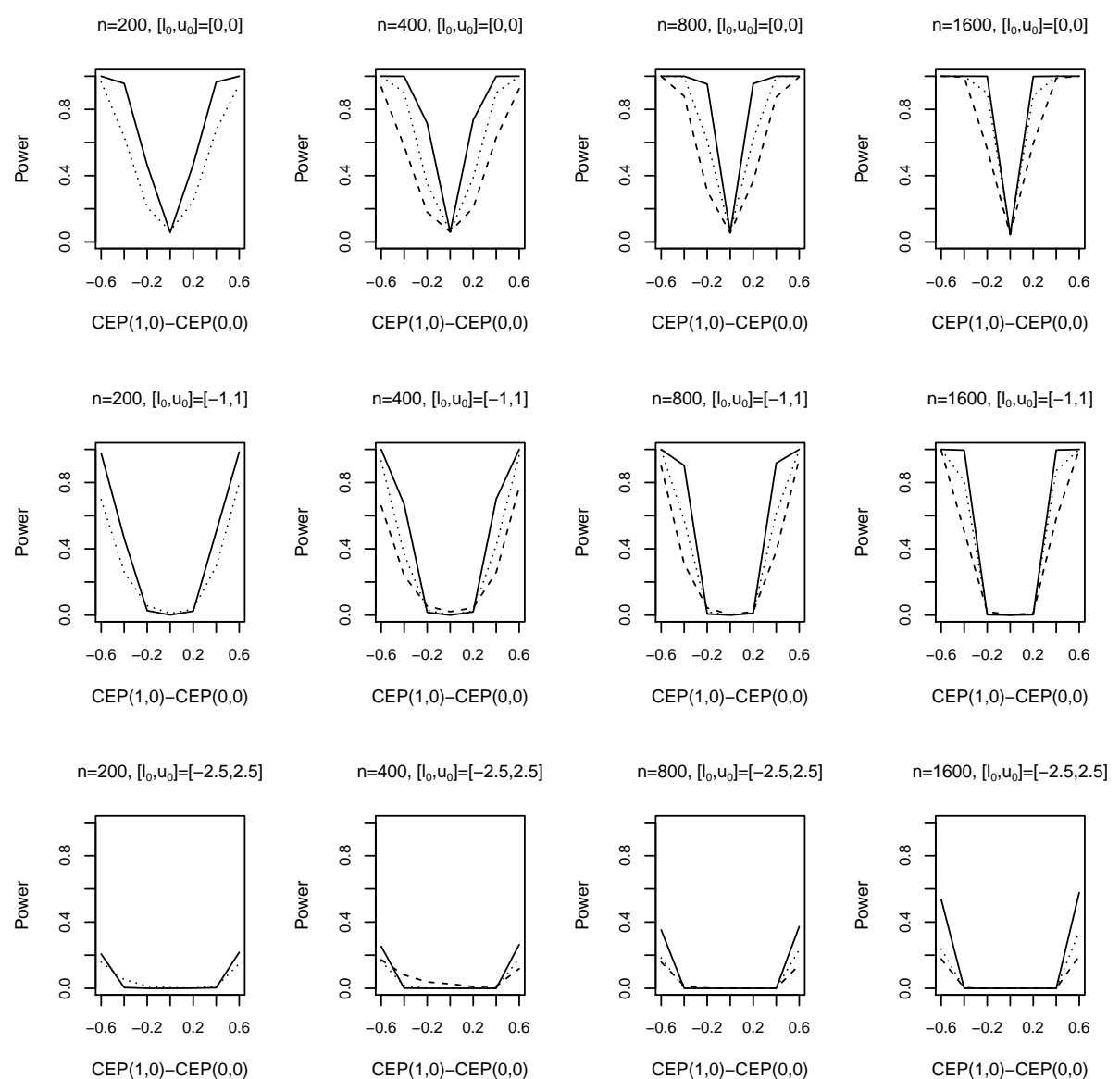


Figure 1: Power to reject $H_0 : \mu = \text{CEP}(1, 0) - \text{CEP}(0, 0)$ for the simulation study under Scenario NEE-CB. Solid black lines denote full cohort and dashed (dotted) lines denote case-cohort with 10% (25%) random subcohort.

set, the data analysis was done using the methods for the scenario **NEE-CB** assumption set. The null hypothesis H_0 was rejected if and only if the 95% EUI for $\mu = CEP(1, 0) - CEP(0, 0)$ calculated as described in Section 5 excluded 0. Power was estimated by the proportion of simulated data sets where H_0 was rejected (Figure 1).

Empirical type I error was close to the nominal level $\alpha = 0.05$. Power increased with sample size and decreased as the interval $[l_0, u_0]$ became wider and as the subcohort size decreased. Web Figure 1 (Web Appendix D) shows the average widths of the 95% EUIs for μ . The EUIs cover at approximately the nominal rate when $l_0 = u_0$, i.e., when μ is identifiable, and are conservative otherwise. The widths are relatively constant across values of μ and increase as the subcohort size decreases. Empirical coverage of the 95% EUIs are plotted in Web Figure 2. Web Figures 3 and 4 display bias of the $\hat{\mu}$ estimates and ratios of the empirical standard errors (ESE) to the average of the sandwich variance estimated standard errors (ASE), showing unbiasedness of both the point and standard error estimators.

A second simulation study was conducted, with data simulated under scenario **NEH-CB** such that **A4** in scenario **NEE-CB** failed. Web Appendix D describes details and results, with data analysis by the methods under assumption-set **NEH-CB** (Web Figures 5–9).

7 Application: HVTN 505 HIV Vaccine Efficacy Trial

We apply the new PS methods to HVTN 505. The PFS biomarker described in the introduction [2] is a quantitative aggregate score derived from constituent qualitative biomarkers that are of interest for binary intermediate outcome PS analysis. The data from a vaccine recipient's Month 6.5 blood sample are measurements of expression of 6 different functional markers (Granzyme-B, IL4, CD40L, TNFalpha, IL2, IFNgamma) in CD8+ T cells after stimulation with HIV-1 peptides. The Bayesian method COMPASS [35] provided, for each vaccine recipient and each of the $2^6 = 64$ possible cell subsets (i.e., combinations of functional markers), the probability the subset was expressed. A question of interest is whether the vaccine effect on HIV-1 infection varied by expression yes vs. no for any of the 64 subsets. Identification of specific cell subsets associated with vaccine protection or harm would give clues on cellular mechanisms of vaccine effect.

In this application Y^τ is the indicator of HIV-1 infection diagnosis between enrollment and the Month $\tau = 6.5$ study visit, and Y is the indicator of this event by the Month 24 final follow-up visit; moreover S^τ is a binary biomarker measured from a Month 6.5 visit blood sample. We defined a vaccine recipient as expressing a given T cell subset ($S^\tau = 1$) if the COMPASS posterior probability exceeded 0.9, and as not expressing the subset ($S^\tau = 0$) if the posterior probability was below 0.1. We restricted to cell subsets with at least 25% of vaccine recipients expressing and at least 25% not expressing the subset, to focus on subsets with ample data support. This yielded three cell subsets for analysis, defined by (GranzymeB, IL4, CD40L, TNFalpha, IL2, IFNgamma) expression pattern of (1) $(-, -, -, -, +, +)$, (2) $(+, -, -, -, +, +)$, and (3) $(+, -, +, +, -, +)$, where + or - indicate that cells do or do not have the function. Among uninfected vaccine recipients (i.e., sampled controls with $Y = 0$) with S^τ data, 71 of 110 (64.5%), 70 of 124 (56.5%), and 43 of 125 (34.4%) have $S^\tau = 1$, respectively. Among infected vaccine recipients (i.e., cases with $Y = 1$) with S^τ data, 5 of 21 (23.8%), 3 of 25 (12.0%), and 1 of 25 (4.0%) have $S^\tau = 1$. Given Case CB holds, for each biomarker our goal is inference on $VE(0) = CEP(0, 0)$ and $VE(1) = CEP(1, 0)$, using the vaccine efficacy contrast $h(x, y) = 1 - x/y$ (Table 1). For reasons given below, we implement the methods using the **NEE-CB** method and the **NEB-CB** method. We consider the meaning and plausibility for HVTN 505 of the critical assumptions in question needed for these two approaches – (**A4**, **A6**) and (**A4''**, **A7'**, **A8**), respectively, where **A6**, **A7'**, and **A8** are needed for valid inference as described in Section 5.2. **A4** (**NEE**) states that the vaccine has no individual-level effects on infection between enrollment and Month 6.5, which is defensible given that $\hat{P}(Y^\tau(1) = 1) = 14/1251 = 0.011$ and $\hat{P}(Y^\tau(0) = 1) = 10/1245 = 0.0080$ with Fisher's exact test 2-sided p-value of $p = 0.54$ for a difference. Because $0.011 > 0.0080$, we also consider the **NEB-CB** method that relaxes **A4** to **A4''** (no individual-level beneficial vaccine effects through Month 6.5).

A6 states that the biomarker has a higher frequency of response for at-risk vaccine than placebo recipients, which essentially always holds. Under Case CB as in HVTN 505 (with $P(S^\tau(0) = 1 | Y^\tau = 0) = 0$), it

obviously holds. **A7'** states that there is a harmful vaccine effect (in expectations) through Month 6.5, which is consistent with the point estimates 0.011 vs. 0.0080 but is not supported by the hypothesis testing with $p = 0.54$. In rare event studies such as HVTN 505, **A8** states that the fraction of vaccine recipients with positive biomarker response $S^T = 1$ is not near zero, which holds. In sum, the method for scenario **NEE-CB** may be the most reasonable because it does not require Early Harm **A7'**, and we focus on its results; we also apply the **NEB-CB** method for comparison. The **NEE-CB** and **NEB-CB** methods are implemented as described in Sections 5.3 and 5.5, respectively.

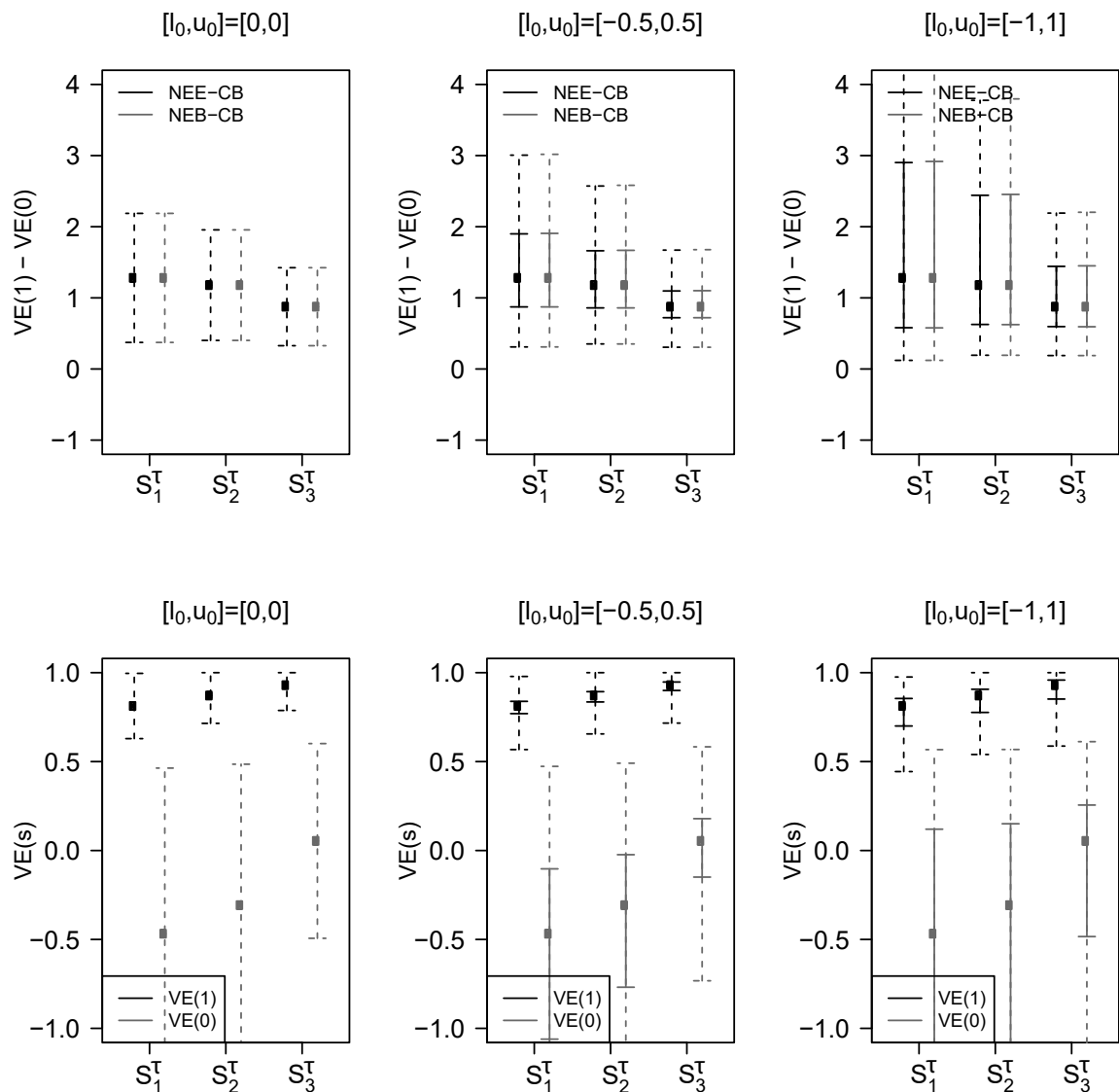


Figure 2: For HVTN 505, ignorance intervals (solid lines) and 95% EUIs (dashed lines) for $VE(0) = CEP(0, 0)$ and $VE(1) = CEP(1, 0)$ and $\mu = VE(1) - VE(0)$ with binary Month 6.5 biomarker S^T equal to 1 (0) if the vaccine recipient expresses (does not express) the CD8+ T-cell subset defined by expression pattern (S_1^T) $(-, -, -, -, +, +)$, (S_2^T) $(+, -, -, -, +, +)$, or (S_3^T) $(+, -, +, +, -, +)$. The sensitivity analysis allows β_0 and β_1 to vary over $[l_0, u_0] = [0, 0]$, $[-0.5, 0.5]$, or $[-1, 1]$; circles are point estimates assuming no selection bias. The top panel shows results for $\mu = VE(1) - VE(0)$ with the black and grey lines (left and right) results for the scenario **NEE-CB** and **NEB-CB** method, respectively. The bottom panel shows scenario **NEE-CB** results for $VE(1)$ and $VE(0)$.

Figure 2 shows the results in terms of ignorance intervals and 95% EUIs under each of the three ranges of sensitivity parameters specified in the simulations. To interpret the sensitivity analysis, note that the single sensitivity parameter β_0 for the **NEE-CB** method has interpretation as the log odds ratio that the biomarker response to vaccination $S^\tau(1)$ takes value 0 for a placebo recipient who is diagnosed with HIV-1 infection between Month 6.5 and 24 compared to a placebo recipient who is not diagnosed with HIV-1 infection through 24 months. Setting β_0 to 0, 0.5, and 1 specifies no, intermediate, and highest degree of selection bias, corresponding to the odds ratio e^{β_0} varying over [1,1], [0.61,1.65], and [0.37,2.72]. The **NEB-CB** method uses the additional sensitivity parameter β_5 (defined in Section 5.5), which is the log odds ratio of infection for a placebo recipient, comparing the always early uninfected (AEU) subgroup (numerator) to the early harmed subgroup (denominator); thus it measures the degree to which vaccine-caused infection by Month 6.5 is prognostic for infection by Month 24 if assigned placebo. It was varied over the same range as β_0 .

On the results, first note that the **NEE-CB** and **NEB-CB** methods give very similar results (top panel), with EUIs for the latter method only very slightly wider (and thus the bottom panel only shows results from the **NEE-CB** method). Web Appendix E (especially Web Figure 10) studies why the methods under scenarios **NEE-CB** and **NEB-CB** are essentially equivalent for HVTN 505, with extra simulations suggesting that a key part of the explanation is that $P(Y^\tau(1) = 1, Y^\tau(0) = 0)$ is small. Second, the results are similar for the three markers, with evidence for effect modification across the expressed vs. not expressed subgroups based on the 95% EUIs for $\mu = VE(1) - VE(0)$ lying above zero even when allowing for the largest amount of potential selection bias ($\beta = 1$, Figure 2 upper-right panel, with lower EUI limit 0.12, 0.19, 0.18 for marker $S_1^\tau, S_2^\tau, S_3^\tau$).

The bottom panel shows that $VE(1)$ is estimated to be positive for each marker (ignorance intervals 0.70–0.85, 0.78–0.91, 0.85–0.96 and 95% EUIs 0.44–0.98, 0.54–1.0, 0.59–0.96, respectively), whereas $VE(0)$ is estimated to be less than zero with wide estimated uncertainty intervals. Based on ignorance intervals these results support a qualitative interaction for each marker. However, because the upper 95% EUI limits for $VE(0)$ extend well above 0, there is not compelling evidence for qualitative interactions.

Remarkably, the results support differential vaccine efficacy according to whether the vaccine induced CD8+ T cells with expression vs. not expression of specific cell subsets, suggesting that the vaccine may have conferred partial protection for individuals with certain identified expression signatures. This motivates follow-on basic science studies of cellular mechanisms of protection. The HIV vaccine field has “moved on” from the DNA/rAd5 type of HIV vaccine platform, no longer considering it. Thus these new results sound a note of caution to not prematurely abandon this platform, in suggesting that if a new version of the regimen could be invented that induces the specific cell subset responses in a much larger subgroup of vaccine recipients, then it could potentially confer high enough overall vaccine efficacy to confer worthwhile public health benefit.

8 Discussion

A sizable literature on nonparametric and semiparametric methods for inference on the survivor average causal effect (SACE) parameter has developed over the past 20 years. Motivated by the need for more robust principal surrogate (PS) analysis in HVTN 505, we described how these methods can be adapted to PS analysis for a binary intermediate variable. This provides new methods for PS analysis, with novel features summarized in the Introduction, all of which were needed for the HVTN 505 application, to appropriately account for the study design and available data and to relax questionable and untestable assumptions. The new methods have application to similar PS questions in other randomized clinical trials.

More robust PS assessment of a qualitative interaction in HVTN 505 was important given the appropriate skepticism that the DNA/rAd5 vaccine could have both beneficial and detrimental effects. The data analysis supported that DNA/rAd5 vaccinated subgroups defined by induction of CD8+ T cells with specific functions had beneficial vaccine efficacy, motivating further research to re-engineer the vaccine regimen to increase induction of these cells.

PS analysis can be improved by accounting for baseline covariates associated with principal strata and/or the final outcome [16, 36]. The specific PS methods described in Section 5 can account for baseline covariates by implementing any preferred covariate-adjusted estimator of the means $risk_1, risk_0, p(0, 0), p(1, 0), p(1, 1)$. An appealing alternative approach that fully accounts for baseline covariates – principal score methods [10–12] – provide PS analysis under a different assumption, principal ignorability, and provide an alternative sensitivity analysis. In general, given that the key steps in estimating the $CEP(s_1, s_0)$ parameters is estimation of the three SACES listed in Table 2, any SACE method that accounts for baseline covariates can be integrated into a new method of PS analysis.

Acknowledgement: We thank the participants and investigators of HVTN 505. Research reported in this publication was supported by the National Institute Of Allergy And Infectious Diseases of the National Institutes of Health under Award Number R37AI054165 and the U.S. Public Health Service Grant AI068635. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

- [1] S. Hammer, M. Sobieszczyk, H. Janes, S. Karuna, M. J. Mulligan, D. Grove, B. Koblin, and et al., Efficacy trial of a DNA/rAd5 HIV-1 preventive vaccine, *N. Engl. J. Med.* 369 (2013), no. 22, 283–292.
- [2] H. E. Janes, K. W. Cohen, N. Frahm, S. C. De Rosa, B. Sanchez, J. Hural, C. A. Magaret, and et al., Higher T-cell responses induced by DNA/rAd5 HIV-1 preventive vaccine are associated with lower HIV-1 infection risk in an efficacy trial, *J. Infect. Dis* 215 (2017), no. 9, 1376–1385.
- [3] C. Frangakis and D. Rubin, Principal stratification in causal inference, *Biometrics* 58 (2002), no. 1, 21–29.
- [4] D. Follmann, Augmented designs to assess immune response in vaccine trials, *Biometrics* 62 (2006), no. 4, 1161–1169.
- [5] C. E. Frangakis and D. B. Rubin, Addressing complications of intention-to-treat analysis in the combined presence of all-or-none treatment-noncompliance and subsequent missing outcomes, *Biometrika* 86 (1999), no. 2, 365–379.
- [6] D. Rubin, Comment on “Causal inference without counterfactuals,” by A.P. Dawid, *J. Am. Stat. Assoc.* 95 (2000), no. 450, 435–437.
- [7] P. B. Gilbert and M. Hudgens, Evaluating candidate principal surrogate endpoints, *Biometrics* 64 (2008), no. 4, 1146–1154.
- [8] E. E. Gabriel, M. C. Sachs, and P. B. Gilbert, Comparing and combining biomarkers as principal surrogates for time-to-event clinical endpoints, *Stat. Med.* 34 (2015), no. 3, 381–395.
- [9] Y. Huang, P. B. Gilbert, and J. Wolfson, Design and estimation for evaluating principal surrogate markers in vaccine trials, *Biometrics* 69 (2013), no. 2, 301–309.
- [10] B. Jo and E. A. Stuart, On the use of propensity scores in principal causal effect estimation, *Stat. Med.* 28 (2009), no. 23, 2857–2875.
- [11] E. A. Stuart and B. Jo, Assessing the sensitivity of methods for estimating principal causal effects, *Stat. Meth. Med. Res.* 24 (2015), no. 6, 657–674.
- [12] P. Ding and J. Lu, Principal stratification analysis using principal scores, *J. R. Stat. Soc. Ser. B Methodol.* 79 (2017), no. 3, 757–777.
- [13] Y. Wang and J. Taylor, A measure of the proportion of treatment effect explained by a surrogate marker, *Biometrics* 58 (2002), no. 4, 803–812.
- [14] Y. Li, J. Taylor, and M. Elliott, A Bayesian approach to surrogacy assessment using principal stratification in clinical trials, *Biometrics* 66 (2010), no. 2, 523–531.
- [15] C. Zigler and T. Belin, A Bayesian approach to improved estimation of causal effect predictiveness for a principal surrogate endpoint, *Biometrics* 68 (2012), no. 3, 922–932.
- [16] D. B. Rubin, Causal inference through potential outcomes and principal stratification: application to studies with “censoring” due to death, *Stat. Sci.* 21 (2006), no. 3, 299–309.
- [17] J. Zhang and D. Rubin, Estimation of causal effects via principal stratification when some outcomes are truncated by “death”, *J. Educ. Behav. Stat.* 28 (2003), no. 4, 353–368.
- [18] M. Hudgens, A. Hoering, and S. Self, On the analysis of viral load endpoints in HIV vaccine trials, *Stat. Med.* 22 (2003), no. 14, 2281–2298.
- [19] K. Imai, Sharp bounds on the causal effects in randomized experiments with ‘truncation by death’, *Stat. Probabil. Lett.* 78 (2008), no. 2, 144–149.
- [20] Y. Chiba, The large sample bounds on the principal strata effect with application to a prostate cancer prevention trial, *Int. J. Biostat.* 8 (2012), no. 1, Article 12.

- [21] D. M. Long and M. G. Hudgens, Comparing competing risk outcomes within principal strata, with application to studies of mother-to-child transmission of HIV, *Stat. Med.* 31 (2012), no. 27, 3406–3418.
- [22] P. B. Gilbert, R. Bosch, and M. Hudgens, Sensitivity analysis for the assessment of vaccine effects on viral load in HIV vaccine trials, *Biometrics* 59 (2003), no. 3, 531–541.
- [23] Y. Jemai, A. Rotnitzky, B. Shepherd, and P. B. Gilbert, Semiparametric estimation of treatment effects given base-line covariates on an outcome measured after a postrandomization event occurs, *J. R. Stat. Soc. Ser. B Methodol.* 69 (2007), no. 5, 879–902.
- [24] B. Shepherd, P. B. Gilbert, and C. Dupont, Sensitivity analyses for comparing time-to-event outcomes only existing in a subset selected postrandomization and relaxing monotonicity, *Biometrics* 67 (2011), no. 3, 1100–1110.
- [25] Y. Chiba and T. vanderWeele, A simple method for principal strata effects when the outcome has been truncated due to death, *Am. J. Epidemiol.* 173 (2011), no. 7, 745–751.
- [26] E. Tchetgen Tchetgen, Identification and estimation of survivor average causal effects, *Stat. Med.* 33 (2014), no. 21, 3601–3628.
- [27] L. Wang, X.-H. Zhou, and T. S. Richardson, Identification and estimation of causal effects with outcomes truncated by death, *Biometrika* 104 (2017), no. 3, 597–612.
- [28] M. Zhang, A. Tsiatis, and M. Davidian, Improving efficiency of inferences in randomized clinical trials using auxiliary covariates, *Biometrics* 64 (2008), no. 3, 707–715.
- [29] J. Robins, A. Rotnitzky, and L. Zhao, Estimation of regression-coefficients when some regressors are not always observed, *J. Am. Stat. Assoc.* 89 (1994), no. 427, 846–866.
- [30] S. Rose and M. J. van der Laan, A targeted maximum likelihood estimator for two-stage designs, *Int. J. Biostat* 7 (2011), no. 1, 1–21.
- [31] B. Saul and M. Hudgens, The calculus of M-estimation in R with geex, arXiv:1709.01413, (2017), [stat.AP] [Preprint] 5 Sep 2017. Last revised 5 Jan 2019. Cited 3 Sep 2019. Available from <https://arxiv.org/abs/1709.01413>.
- [32] G. Imbens and C. Manski, Confidence intervals for partially identified parameters, *Econometrica* 72 (2004), no. 6, 1845–1857.
- [33] S. Vansteelandt, E. Goetghebeur, M. Kenward, and G. Molenberghs, Ignorance and uncertainty regions as inferential tools in a sensitivity analysis, *Stat. Sin.* 16 (2006), 953–979.
- [34] A. Richardson, M. Hudgens, P. B. Gilbert, and J. Fine, Nonparametric bounds and sensitivity analysis of treatment effects, *Stat. Sci.* 29 (2014), no. 4, 596–618.
- [35] L. Lin, G. Finak, K. Ushey, C. Seshadri, T. R. Hawn, N. Frahm, T. J. Scriba, H. Mahomed, W. Hanekom, P.-A. Bart, et al., COMPASS identifies T-cell subsets correlated with clinical outcomes, *Nat. Biotech.* 33 (2015), no. 6, 610.
- [36] D. M. Long and M. G. Hudgens, Sharpening bounds on principal effects with covariates, *Biometrics* 69 (2013), no. 4, 812–819.

Supplementary Materials

Title: Supportive results (A) Identifiability results; (B) Application of Chiba and VanderWeele’s (2011) SACE method for evaluating a binary principal surrogate under Scenarios **NEE-VB** and **NEE-CB**; (C) Extending the methods for a time-to-event outcome with right-censoring; (D) Web Figures 1–9 simulation study results; (E) Additional Application results, including Web Figure 10; (F) R computer code.