# Sensitivity Analysis of Per-Protocol Time-to-Event Treatment Efficacy in Randomized Clinical Trials

**Peter B. Gilbert**[1,*], **Bryan E. Shepherd**[2], **Michael G. Hudgens**[3]

[1] Department of Biostatistics, University of Washington and Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, Washington, 98109, U.S.A.

[2]Department of Biostatistics, Vanderbilt University School of Medicine, Nashville, Tennessee, 37232, U.S.A.

[3]Department of Biostatistics, University of North Carolina, Chapel Hill, North Carolina, 27599, U.S.A.

[*] *e-mail:* pgilbert@scharp.org

SUMMARY. Assessing per-protocol treatment efficacy on a time-to-event endpoint is a common objective of randomized clinical trials. The typical analysis uses the same method employed for the intention-to-treat analysis (e.g., standard survival analysis) applied to the subgroup meeting protocol adherence criteria. However, due to potential post-randomization selection bias, this analysis may mislead about treatment efficacy. Moreover, while there is extensive literature on methods for assessing causal treatment effects in compliers, these methods do not apply to a common class of trials where a) the primary objective compares survival curves, b) it is inconceivable to assign participants to be adherent and event-free before adherence is measured, and c) the exclusion restriction assumption fails to hold. HIV vaccine efficacy trials including the recent RV144 trial exemplify this class, because many primary endpoints (e.g., HIV infections) occur before adherence is measured, and nonadherent subjects who receive some of the planned immunizations may be partially protected. Therefore, we develop methods for assessing per-protocol treatment efficacy for this problem class, considering three causal estimands of interest. Because these estimands are not identifiable

from the observable data, we develop nonparametric bounds and semiparametric sensitivity analysis methods that yield estimated ignorance and uncertainty intervals. The methods are applied to RV144.

KEY WORDS: As-treated; Bounds; Causal inference; Exclusion restriction; Ignorance region; Intention to treat; Principal stratification; Selection bias; Survival analysis.

## 1. Introduction

Over the past 30 years, millions of individuals have acquired HIV and the global rate of new infections remains high. Unfortunately, the development of an HIV vaccine has been difficult with clinical trials not yielding promising results. However, a recent community-based, individually-randomized, multicenter, double-blind, placebo-controlled clinical trial of 16,395 HIV negative volunteers in Thailand (the RV144 'Thai trial') supported that an HIV vaccine regimen had partial efficacy to reduce the risk of HIV infection (Rerks-Ngarm et al. 2009). This trial administered four injections of a recombinant canarypox vector vaccine (or placebo) at the Week 0, 4, 12, 24 study visits, plus two injections of a recombinant glycoprotein 120 subunit vaccine (or placebo) at the Week 12, 24 study visits, and monitored participants for the primary endpoint of HIV infection from entry until the final study visit at 3.5 years. While the Thai trial's finding of partial efficacy generated great enthusiasm as the first positive finding from an HIV vaccine efficacy trial, it also generated confusion, because the results on vaccine efficacy appeared to differ depending on whether the analysis utilized the modified intention-to-treat (MITT) or per-protocol (PP) cohort, where the MITT cohort was all randomized subjects HIV negative at baseline (determined via blinded procedures), and the PP cohort was the subset of the MITT cohort that tested HIV negative at the Week 24 study visit and received all 6 study injections at the Week 0, 4, 12, 24 study visits within pre-specified allowable visit windows. With vaccine efficacy ($VE$) defined as the percent reduction in the cumulative probability of HIV infection diagnosis by

39 months post-randomization (vaccine versus placebo), estimated based on Kaplan-Meier estimates, the MITT result was $\widehat{VE} = 31\%$, 95% CI 0% to 51%, p = 0.04 and the PP result was $\widehat{VE} = 25\%$, 95% CI -16% to 51%, p = 0.19. While the point estimates were fairly close to one another and hence there was little evidence for different effects, nonetheless many observers expressed concern that $VE$ was apparently lower in the PP cohort.

Due to randomization, the comparator groups in the MITT analysis are guaranteed to have balanced baseline prognostic factors on average, such that the MITT analysis provides consistent estimation of a causally-interpretable $VE$ parameter, namely the percent reduction in the cumulative risk of infection if everyone were assigned vaccine compared to if everyone were assigned placebo. On the other hand, the comparator groups in the PP analysis are subsets of randomized subjects whose membership is determined conditional on post-randomization variables, resulting in possible selection bias (Rosenbaum, 1984), such that the PP analysis does not necessarily assess a causal effect of vaccination. Rather, it assesses a mixture of the effect of vaccine assignment and imbalanced prognostic factors created by conditioning on qualification into the PP cohort. The PP analysis included 24% fewer subjects and 31% fewer primary endpoint events than the MITT analysis. Table 1 summarizes the reasons why MITT subjects failed to qualify for the PP cohort. The rate of nonadherence to study injections was significantly higher in the vaccine than placebo group (24% versus 21%, Chi-squared test $p < 0.001$), possibly due to a higher rate of reactogenicity in the vaccine group (79% versus 59%) (Rerks-Ngarm et al. 2009).

As discussed in Gilbert et al. (2011), opinions varied on the interpretation of the results, ranging from discounting the PP analysis because only the MITT analysis provides an asymptotically unbiased answer to a question of clear interest (typically statisticians), to discounting any inference about positive vaccine efficacy because one would expect the vaccine to work better in those who received all of the immunizations

Table 1: Culling of the MITT Cohort to Form the PP Cohort in the RV144 Thai Trial*

| Reason for Exclusion from the PP Cohort | MITT Vaccine | MITT Placebo |
|---|---|---|
| Diagnosed with HIV by the week 24 visit | 5/8197 (0.06%) | 10/8198 (0.12%) |
| Not diagnosed with HIV by the week 24 visit but nonadherent ($A = 0$) | | |
| Dropped out before week 24 visit | 67/8197 (0.82%) | 70/8198 (0.85%) |
| Reached week 24 visit HIV negative but nonadherent to vaccination visits: | 1949/8197 (23.8%) | 1752/8198 (21.4%) |
| All 4 with $\geq 1$ outside window | 749 | 671 |
| Received 3 vaccinations | 80 | 68 |
| Received 2 vaccinations | 81 | 76 |
| Received 1 vaccination | 146 | 134 |
| Received 0 vaccinations | 2 | 4 |
| Other reasons (mainly outside window) | 891 | 799 |
| Total Culled Out | 2021/8197 (24.7%) | 1832/8198 (22.3%) |

*Group-imbalances in prognostic factors for HIV infection could arise due to differences (by treatment assignment) in probabilities of any of the events (1) HIV infection diagnosis by the Week 24 visit, (2) dropout by the Week 24 visit, or (3) reaching the Week 24 visit at-risk but nonadherent to the vaccinations.

(typically non-statisticians). In our view, the former interpreters are correct that the problematic interpretation of the PP analysis renders it of marginal value, whereas the latter interpreters are correct that, were $VE$ in vaccine-adherent subjects to be assessed in a more meaningful way, it would indeed add value.

To improve upon the standard analysis of $VE$ in the PP cohort, an analytic method that adjusts for subject factors known to predict PP cohort membership and HIV infection (such factors may cause selection bias) should be applied (e.g., see Tsiatis et al. 2008), which in addition to correcting for bias can improve statistical power by leveraging prognostic factors. Moreover, because some biasing factors may be unmeasured, the sensitivity of results to such factors should also be investigated. However, existing approaches to sensitivity analysis do not directly apply to vaccine efficacy trials, or, more generally, to two-arm randomized trials of treatments where the primary objective compares survival curves, some participants experience the primary endpoint before adherence is evaluated, and some subjects have partial adherence (implying

that the exclusion restriction assumption is implausible; see Section 3 for the definition and interpretation of this assumption in our context). Therefore in this article we develop methods for this class of trials. The methods apply to most vaccine efficacy trials, given that almost all such trials assess per-protocol $VE$ (Horne, Lachenbruch, and Goldenthal, 2001), and typically multiple immunizations are administered over a period of months during which disease events occur. Moreover, over the past several decades the standard non-causal per-protocol analysis has been ubiquitously applied to such efficacy trials, and this work aims to help improve the standard approach by adding causal assessment.

This manuscript is organized as follows. In Section 2 we define notation and three PP treatment effect estimands, and place the current work in the context of previously developed methodology for evaluating causal effects in compliers. In Section 3 assumptions are given that help identify the estimands under four different assumption sets. In Sections 4 and 5 we describe non- and semi-parametric methods for inference including techniques for estimating ignorance and uncertainty intervals. The methods are developed under multiple assumption sets so that practitioners may tailor the sensitivity analysis approach to their particular trial; where stronger assumptions are warranted the estimated ignorance and uncertainty intervals will tend to be narrower. In Section 6 we apply the methods to the Thai trial. The methods are implemented using the *sensitivityPStrat* package in R available on CRAN and our complete analysis code is posted at the second author's website.

## 2. Notation, Per-Protocol Estimands, and Problem Context

Consider a two-group randomized trial, wherein subjects are randomly assigned to treatment $Z = 1$ or $Z = 0$ (in our motivating example $Z = 1$ is vaccine and $Z = 0$ is placebo) and are followed for the primary time-to-event endpoint (typically a disease event). Let $T$ be the time from randomization until the disease event and $C$ be the

censoring time, such that $X = \min(T, C)$ and $\Delta = I(T \leq C)$ are observed. Let $A$ be an indicator of adherence/compliance to assigned treatment defined based on study visits between randomization through time $\tau_0$. For example, in the Thai trial, study injections were scheduled at Week 0, 4, 12, 24 visits, and $A = 1$ was defined as receipt of all planned injections within the pre-specified allowable visit windows. Note that $A = 1$ for $Z = 1$ subjects indicates appropriate receipt of vaccination injections whereas $A = 1$ for $Z = 0$ subjects indicates appropriate receipt of placebo injections. We treat $A$ as undefined if $T \leq \tau_0$, denoted by $A = *$, and assume the data are iid observations $O_i = (Z_i, X_i, \Delta_i, A_i)$, $i = 1, \cdots, n$.

The population of interest for inference is disease-free and adherent up to $\tau_0$. Thus, the PP cohort of interest is $\{i : T_i > \tau_0, A_i = 1\}$, and we define $PP = I(T > \tau_0, A = 1)$. Let $S_z^{PPobs}(t) \equiv P(T > t | Z = z, T > \tau_0, A = 1)$ be the survival function in the PP cohort if assigned treatment $Z = z$ for $t \geq \tau_0$, and $F_z^{PPobs}(t) \equiv 1 - S_z^{PPobs}(t)$. Also let $S_z^{\tau_0 obs}(t) \equiv P(T > t | Z = z, T > \tau_0)$ and $F_z^{\tau_0 obs}(t) \equiv 1 - S_z^{\tau_0 obs}(t)$. It is widespread in clinical trials to measure per-protocol treatment efficacy by a contrast in $S_1^{PPobs}(t)$ and $S_0^{PPobs}(t)$, which does not measure a causal effect. Below we define three survival causal effect ($SCE$) estimands that account for adherence/compliance.

The first estimand of interest is based on the literature for assessing causal treatment effects in "always compliers," "always infected," or "always survivors" (Angrist, Imbens, and Rubin, 1996; Frangakis and Rubin, 2002; Shepherd, Gilbert, and Lumley, 2007; Shepherd, Gilbert, and Dupont, 2011, henceforth SGD). Let $T(z), C(z), X(z)$, $\Delta(z), A(z)$ be the potential outcomes of $T, C, X, \Delta, A$ under assignment $Z = z$ for $z = 0, 1$. Let $S_z(t) \equiv P(T(z) > t)$, $F_z(t) \equiv 1 - S_z(t)$, and $PP(z) \equiv I(T(z) > \tau_0, A(z) = 1)$, and set $pp_z \equiv P(PP(z) = 1)$. For $x, y \in [0, 1]$ let $h(x, y)$ be a known contrast function satisfying $h(x, x) = 0$ and $h(x_1, y_1) \leq h(x_2, y_2)$ for all $x_1 \leq x_2$ and $y_1 \geq y_2$, such as $h(x, y) = x - y$ or $1 - (1 - x)/(1 - y)$.

With the always-compliers approach, the survival causal effect estimand is

$$SCE^{APP}(t) \equiv h(S_1^{APP}(t), S_0^{APP}(t)) \qquad \text{for } t \geq \tau_0, \tag{1}$$

where $S_z^{APP}(t) \equiv P(T(z) > t | PP(1) = PP(0) = 1)$ for $z = 0, 1$. The estimand $SCE^{APP}(t)$ measures the causal treatment effect in "always per-protocol" ($APP$) subjects, $\{PP(1) = PP(0) = 1\}$, i.e., subjects who would survive to $\tau_0$ under both treatments and would be compliant to whatever they were assigned. One appeal of this estimand is that always compliers may be persuaded to follow a treatment policy.

The second survival causal effect estimand of interest is

$$SCE^{ASA1}(t) \equiv h(S_1^{ASA1}(t), S_0^{ASA1}(t)) \qquad \text{for } t \geq \tau_0, \tag{2}$$

where $S_z^{ASA1}(t) \equiv P(T(z) > t | T(1) > \tau_0, T(0) > \tau_0, A(1) = 1)$ for $z = 0, 1$, which evaluates "always survivors" (AS) to time $\tau_0$ who would be adherent under $Z = 1$. This estimand differs from $SCE^{APP}(t)$ by measuring treatment efficacy in those adherent to treatment 1, regardless of whether they would adhere to treatment 0. This subpopulation is of interest because, once a treatment is licensed, treatment decisions and policies will largely be based on the predicted effect of that treatment for those who take it, whereas adherence to a control preparation may no longer be relevant. Follmann (2000) and Loeys and colleagues cited below studied the same subpopulation of "would-be treatment compliers." We view the $APP$ and $ASA1$ estimands as complementary, each of interest in its own right and capturing different information.

While there is a rich literature for assessing causal treatment effects in always compliers or would-be treatment compliers, most papers have not accommodated an outcome subject to censoring, and, among those that do, an assumption is made that renders the method inapplicable in our context. Cuzick et al. (2007) developed a structural proportional hazards model (SPHM) for estimating a hazard ratio estimand in the $APP$ subpopulation; however, they assume adherence status $A$ is known at randomization. Closer to our setting in assuming $A$ is measured after randomization,

Loeys and Goetghebeur (2003) and Loeys, Goetghebeur, and Vandebosch (2005) developed an SPHM for estimating a hazard ratio estimand in the $ASA1$ subpopulation assuming all-or-none or time-constant adherence and $A(0) = 1$ for all subjects, in which case the method equivalently applies to the $APP$ subpopulation. This work does not apply to our setting because it assumes the exclusion restriction, and because we are specifically interested in comparing survival curves, which allows assessing treatment efficacy over time and avoids the built-in selection bias of hazard ratios (Hernán, 2010). Baker (1998) and Nie, Cheng, and Small (2011) developed nonparametric likelihood estimation-based approaches for the $APP$ estimand, but, in requiring the exclusion restriction, they also do not apply. SGD provided an applicable approach for the $APP$ estimand (1) that does assume the exclusion restriction, and a novel contribution of this article is to extend SGD to new identifiability assumption sets that are sometimes reasonable in double-blinded randomized trials.

The third survival causal effect estimand we study is

$$SCE^{PP1}(t) \equiv h(S_1^{PP1}(t), S_0^{PP1}(t)) \qquad \text{for} \quad t \geq \tau_0, \tag{3}$$

where $S_z^{PP1}(t) \equiv P(T(z) > t | PP(1) = 1)$ for $z = 0, 1$, which evaluates the PP cohort under treatment 1. This estimand addresses the question: "What is the treatment efficacy for the subpopulation that would satisfy the PP-criteria if assigned treatment 1?" We are unaware of statistical methods for this estimand, perhaps in part because the estimand in a sense favors treatment 1, as by definition $S_1^{PP1}(\tau_0) = 1$, and hence treatment 1 starts out at least as good as treatment 0. Moreover, when $T(1)$ and $T(0)$ are not perfectly correlated, it is easy to find examples where there is no marginal causal treatment effect on $T$, but there is an expected beneficial treatment effect $S_1^{PP1}(t) > S_0^{PP1}(t)$ for all $t > \tau_0$ (see Supplementary Materials, Section 1). Despite this issue, we include this estimand because the per-protocol cohort under study can be directly observed for the treatment 1 group, and it addresses a question of interest for treatment

1 PP subjects: "Compared to their observable survival experience, what would their survival experience have been had they been assigned treatment 0?"

## 3. Assumption Sets

Throughout we make the "base set" of assumptions made for randomized clinical trials: Stable Unit Treatment Values (SUTVA), Ignorable Treatment Assignment ($Z \perp T(1), T(0), C(1), C(0), A(1), A(0)$), and Random Censoring ($T(z) \perp C(z)$ for $z = 0, 1$), where $\perp$ denotes independence. SUTVA states that potential outcomes for each subject $i$ are unrelated to the assignment $Z_j$ of other subjects and that there are not multiple versions of treatment. We also consider additional assumptions involving the initial study period through $\tau_0$, which will be utilized to help identify the estimands.

**Survival Monotonicity (SM):** $P(T(1) \leq \tau_0 < T(0)) = 0$

**Adherence Monotonicity (AM):** $P(A(1) = 1, A(0) = 0 | T(1) > \tau_0, T(0) > \tau_0) = 0$

**Equal Adherence (EA):** $P(A(1) = A(0) | T(1) > \tau_0, T(0) > \tau_0) = 1$

SM states that individuals who would survive beyond $\tau_0$ if assigned $Z = 0$ would also survive beyond $\tau_0$ if assigned $Z = 1$. Examples of trials where SM may hold are placebo-controlled trials and non-inferiority trials. AM states that for individuals surviving beyond $\tau_0$ under both assignments, those who would adhere under assignment $Z = 1$ would also adhere under assignment $Z = 0$. Adherence requires retention and receipt of treatment through $\tau_0$, such that AM assumes that all subjects retained and adherent through $\tau_0$ under $Z = 1$ would also be retained and adherent through $\tau_0$ under $Z = 0$. For a double-blinded trial, differential drop-out and adherence would likely be due to differential side effects, such that AM is plausible if side effects are as or more likely for $Z = 1$ than for $Z = 0$. Here the double-blinding may be crucial, as it implies there should not be less adherence in the control group due to a preference to receive treatment. The stronger assumption EA states that all subjects retained and adherent

under one assignment would also be so under the other. While this is implausible for trials of treatments with serious side effects, it may be plausible for double-blinded trials of non-toxic interventions in healthy populations (e.g., prevention trials).

We consider four types of assumption sets under which sensitivity analysis techniques are developed: **(A)** base set; **(B)** base set plus AM; **(C)** base set plus (SM, AM); **(D)** base set plus (SM, EA). These sets posit increasingly stringent assumptions. SM, AM, and EA can be rejected based on data (e.g., SM is rejected if a test indicates $S_1(\tau_0) < S_0(\tau_0)$ in any baseline subgroup), but they cannot be fully verified, given that they must hold for every individual. Supplementary Table 1 describes the basic principal strata (Frangakis and Rubin, 2002) formed based on $T(1), T(0), A(1), A(0)$, as well as the subpopulations defining the estimands under the assumption sets.

Given the relatedness of our problem with causal inference for compliers, it is noteworthy that we do not consider the exclusion restriction assumption that is commonly made. In our context this assumption is expressed as $P(T(1) = T(0)|T(1) > \tau_0, T(0) > \tau_0, A(1) = A(0) = 0) = 1$, i.e., the treatment has no effect in individuals non-adherent under each treatment assignment. This assumption may be plausible in settings with all-or-none adherence, for which $A(1) = 0$ indicates no receipt of treatment whatsoever. However, in vaccine trials, several immunizations are planned and many subjects with $A(1) = 0$ receive some or even all immunizations (as for the Thai trial, see Table 1), and hence could have a beneficial vaccine effect on infection (i.e., $T(1) > T(0)$). Also, unlike SM, AM, EA, the exclusion restriction would require making an assumption about the time period after $\tau_0$.

## 4. Semiparametric Sensitivity Analysis and Nonparametric Bounds

We develop semiparametric modeling approaches to sensitivity analysis for each estimand under each assumption set, which use a fixed sensitivity parameter(s) $\gamma$ (not identifiable from the observed data and the base set of assumptions) within a specified

region $\Gamma$; accordingly $S_1^{\#}(t,\gamma)$, $S_0^{\#}(t,\gamma)$, and $SCE^{\#}(t;\gamma)$ are indexed by $\gamma \in \Gamma$, where $\#$ denotes $APP$, $ASA1$, or $PP1$. We also develop nonparametric bounds, obtained by setting the sensitivity parameter(s) to extreme values. Our approach follows Robins (1997) and Vansteelandt et al. (2006), where each estimand is nonparametrically identified once the sensitivity parameter(s) $\gamma$ indexing the full data law $\mathcal{M}(\gamma)$ defined by modeling restrictions is fixed, and the goal is inference on the estimand under the union model $\mathcal{M}(\Gamma) = \cup_{\gamma\in\Gamma}\mathcal{M}(\gamma)$, assuming the true value $\gamma_0$ of $\gamma$ lies in $\Gamma$.

**4.1. Estimands $S_1^{APP}(t)$ and $S_0^{APP}(t)$.**

*Assumption set A.* As noted above, we will adapt the general sensitivity analysis method of SGD for making inferences about the $APP$ estimand. The base set of assumptions plus the following three selection models (and the observed data) identify $S_1^{APP}(t)$ and $S_0^{APP}(t)$:

**B.1:** $P(PP(1) = 1|PP(0) = 1, T(0) = t) = w_0(t;\alpha_0,\beta_0)$ for $t > \tau_0$, where $w_0(t;\alpha_0,\beta_0) = G_0\{\alpha_0 + h_0(t;\beta_0)\}$, $\beta_0$ is a fixed and known parameter, $G_0(\cdot)$ is a known cdf, $\alpha_0$ is an unknown parameter, and for each $\beta_0$, $h_0(t;\beta_0)$ is a known function of $t$.

**B.2:** $P(PP(0) = 1|PP(1) = 1, T(1) = t) = w_1(t;\alpha_1,\beta_1)$ for $t > \tau_0$, where $w_1(t;\alpha_1,\beta_1) = G_1\{\alpha_1 + h_1(t;\beta_1)\}$, $\beta_1$ is a fixed and known parameter, $G_1(\cdot)$ is a known cdf, $\alpha_1$ is an unknown parameter, and for each $\beta_1$, $h_1(t;\beta_1)$ is a known function of $t$.

**B.3:** $\pi_{APP} \equiv P(PP(1) = PP(0) = 1)$ is fixed and known.

B.1 models the dependency of $PP(1)$ on the failure time $T(0)$ in the $\{PP = 1, Z = 0\}$ subgroup, and B.2 posits a similar model with the treatment assignments swapped. The parameters $\gamma = (\beta_0, \beta_1, \pi_{APP})$ are not identified by the observed data; they are sensitivity parameters, treated as known in a single analysis and varied over a range of values to form a sensitivity analysis (Scharfstein, Rotnitzky, and Robins, 1999). Following SGD, we specify each $w_z(t;\alpha_z,\beta_z)$ for $z = 0, 1$ with a modified inverse logit

function,

$$w_z(t;\alpha_z,\beta_z) = [1+\exp\{-\alpha_z - \beta_z \min(t,\tau)\}]^{-1}, \tag{4}$$

where $\tau > \tau_0$ is some number less than or equal to the maximum length of follow-up. The minimum of $t$ and $\tau$ is used instead of $t$ to avoid parametric assumptions about $S_1^{APP}(t)$ and $S_0^{APP}(t)$ beyond the support of the data. Based on (4), $\beta_0$ is the log odds ratio of meeting the PP-criteria under treatment $Z = 1$ per unit increment of $T(0)$ for subjects who meet the PP-criteria under treatment $Z = 0$. This parameter is interpreted for per-protocol placebo recipients with different infection times if not vaccinated, $T(0)$, which reflects their risk for infection. As such, $\beta_0 = 0$ implies that the odds of being per-protocol if vaccinated is not associated with infection risk in per-protocol placebo recipients. If $\beta_0 > 0$ ($\beta_0 < 0$), then the odds of being per-protocol if vaccinated are higher among per-protocol placebo recipients at lower (higher) risk for infection, i.e., with longer (shorter) infection times. The parameter $\beta_1$ has a parallel interpretation for per-protocol vaccine recipients.

Models B.1 and B.2 may be equivalently expressed as pattern mixture models, which give the sensitivity parameters alternative interpretations that may be preferred by some practioners. For example, with weight (4), B.1 can be re-expressed as

$$\exp\{\beta_0 t\} = \frac{\Pr(T(0)=t|PP(1)=1,PP(0)=1)}{\Pr(T(0)=t|PP(1)=0,PP(0)=1)} \frac{\pi_{APP}}{\Pr(PP(1)=0,PP(0)=1)}$$

for $t \in (\tau_0, \tau]$, such that $\exp\{\beta_0\}$ is the density ratio (comparing the $APP$ and $\{PP(1)=0,PP(0)=1\}$ subpopulations) of infection at time $t$ divided by this density ratio at time $t-1$. Thus $\beta_0$ calibrates the discrepancy between the distributions of $T(0)$ for the $APP$ and $\{PP(1)=0,PP(0)=1\}$ subpopulations.

B.3 specifies the probability a subject is always per-protocol. Instead of specifying $\pi_{APP}$, alternatively one could specify $\phi_{APP} \equiv \pi_{APP}/pp_1 = P(PP(0)=1|PP(1)=1)$.

Under B.1 and B.2, for $z = 0, 1$

$$P(T(z) \le t|PP(1)=PP(0)=1) = \frac{\int_{\tau_0}^{t} w_z(s;\alpha_z,\beta_z) dF_z^{PPobs}(s)}{\int_{\tau_0}^{\infty} w_z(s;\alpha_z,\beta_z) dF_z^{PPobs}(s)}. \tag{5}$$

Furthermore, algebra shows that

$$\int_{\tau_0}^{\infty} w_z(s;\alpha_z,\beta_z)dF_z^{PPobs}(s) = \pi_{APP}/pp_z, \tag{6}$$

and therefore

$$S_z^{APP}(t) = 1 - \frac{pp_z}{\pi_{APP}} \int_{\tau_0}^{t} w_z(s;\alpha_z,\beta_z)dF_z^{PPobs}(s), \quad \text{for } z = 0,1. \tag{7}$$

Once $\beta_0, \beta_1, \pi_{APP}$ are specified, $\alpha_0$ and $\alpha_1$ are identified based on (6), such that $S_0^{APP}(t)$ and $S_1^{APP}(t)$ are identified.

Under the base set of assumptions alone, nonparametric bounds for $S_z^{APP}(t)$ are achieved by setting $\pi_{APP}$ to its smallest possible value and $\beta_z$ to $-\infty$ and $\infty$, i.e., $w_z(t) = I(t \leq q_z^{\pi_{APP}/pp_z})$ and $w_z(t) = I(t \geq q_z^{1-\pi_{APP}/pp_z})$, where $q_z^a$ is the $a^{\text{th}}$ percentile of $F_z^{PPobs}(\cdot)$, for $z = 0,1$. Define $\pi_{min} \equiv \max\{0, pp_0 + pp_1 - 1\}$, the smallest possible value of $\pi_{APP}$. If $\pi_{min} = 0$, then the bounds for $S_z^{APP}(t)$ are the uninformative values 0 and 1. If $\pi_{min} > 0$, then $S_{z,lo}^{APP}(t) \leq S_z^{APP}(t) \leq S_{z,hi}^{APP}(t)$, where

$$S_{z,lo}^{APP}(t) \equiv \max\left\{0, 1 - \frac{F_z^{PPobs}(t)}{\pi_{min}/pp_z}\right\}; S_{z,hi}^{APP}(t) \equiv \min\left\{\frac{pp_z}{\pi_{min}} - \frac{F_z^{PPobs}(t)}{\pi_{min}/pp_z}, 1\right\}. \tag{8}$$

for $t \geq \tau_0$ and $z = 0,1$. These bounds determine the nonparametric bounds for $SCE^{APP}(t)$ for any contrast function $h(\cdot,\cdot)$ satisfying the properties described in Section 2, equal to

$$h\left(S_{1,lo}^{APP}(t), S_{0,hi}^{APP}(t)\right) \leq SCE^{APP}(t) \leq h\left(S_{1,hi}^{APP}(t), S_{0,lo}^{APP}(t)\right) \quad \text{for } t \geq \tau_0. \tag{9}$$

These bounds are nonparametric because no parametric modeling assumptions are used in expressing the bounds or in the procedures for estimating them (e.g., B.1–B.3 are not needed). The Supplementary Materials (Section 2) contain a proof of (8).

*Assumption set B (add AM).* The same approach is used when AM is added, with the only difference being that the amount of possible post-randomization selection bias is reduced. In particular, $\pi_{APP}$ is now constrained between $\max\{0, S_0(\tau_0) + pp_1 - 1\}$

and $\min\{pp_0, pp_1\}$. In addition, in B.2 the weight function $w_1(t;\alpha_1,\beta_1) = P(A(0) = 1|T(1) > \tau_0, T(0) > \tau_0, A(1) = 1, T(1) = t)P(T(0) > \tau_0|T(1) > \tau_0, A(1) = 1, T(1) = t)$, and by AM the first conditional probability is known to be one.

The nonparametric bounds for the $APP$ estimand under assumption set B are as in (8) and (9) with $\pi_{min}$ modified to $\pi_{min} = \max\{0, S_0(\tau_0) + pp_1 - 1\}$, and are always at least as narrow as those under assumption set A. If there is nonadherence under treatment $z = 0$ such that $pp_0 < S_0(\tau_0)$, then $\pi_{APP}$ has a narrower range via the extra assumption AM, yielding narrower bounds (if $\pi_{min} > 0$). This occurs in the Thai trial example.

*Assumption set C (add SM, AM).* Again the same approach is used as under assumption sets A and B, with the amount of possible post-randomization bias further reduced by the constraint $\max\{0, S_0(\tau_0) + pp_1 - S_1(\tau_0)\} \leq \pi_{APP} \leq \{pp_1, pp_0\}$. Under SM, $PP(1) = A(1)$, which simplifies the interpretation of $\beta_0$ to reflect the association of infection risk of per-protocol placebo recipients with adherence if vaccinated. The nonparametric bounds are the same as above with $\pi_{min}$ replaced with $\pi_{min} = \max\{0, S_0(\tau_0) + pp_1 - S_1(\tau_0)\}$.

*Assumption set D (add SM, EA).* By EA, $w_0(t;\alpha_0,\beta_0)$ in B.1 equals $P(T(1) > \tau_0|T(0) > \tau_0, A(0) = 1, T(0) = t)$, which by SM equals one. Therefore $S_0^{APP}(t) = S_0^{PPobs}(t)$. In addition, by EA model B.2 is expressed as $w_1(t;\alpha_1,\beta_1) = P(T(0) > \tau_0|T(1) > \tau_0, A(1) = 1, T(1) = t)$, which we again model with the modified inverse logit function (4). SM and EA together identify $\pi_{APP}$ as $\pi_{APP} = pp_0$. Therefore only B.2 among the models (B.1, B.2, B.3) is needed, and simple calculations show that

$$S_1^{APP}(t) = 1 - \frac{pp_1}{pp_0}\int_{\tau_0}^{t} w_1(s,\alpha_1,\beta_1)dF_1^{PPobs}(s) \qquad \text{for } t \geq \tau_0, \tag{10}$$

where $\alpha_1$ is the solution of (6) with $z = 1$ and $\pi_{APP} = pp_0$. The nonparametric bounds for $S_1^{APP}(t)$ (for $t \geq \tau_0$) are given by (8) with $\pi_{min} = pp_0$ and $z = 1$.

*Nonparametric bounds under alternative assumption sets.* Zhang and Rubin (2003)

and Chiba (2012) derived nonparametric bounds for the average causal effect in "always survivors" under different assumptions than SM, AM, and EA. While we chose our assumption sets to be maximally relevant for vaccine trials, these alternative assumptions may also be plausible for some vaccine trials, and yield alternative nonparametric bounds for $SCE^{APP}(t)$. Seven alternative assumption sets, their interpretations and plausibility for RV144, and the corresponding nonparametric bounds are described in Supplementary Materials (Section 3 and Supplementary Table 2).

**4.2. Estimands $S_1^{ASA1}(t)$ and $S_0^{ASA1}(t)$.**

*Assumption set A.* The $ASA1$ estimands are identified using the same models as above, except the event $\{PP(0)=1\}$ is replaced with $\{T(0)>\tau_0\}$:

**B.1':** $P(PP(1)=1|T(0)>\tau_0, T(0)=t) = w_0(t;\alpha_0,\beta_0')$.

**B.2':** $P(T(0)>\tau_0|PP(1)=1, T(1)=t) = w_1(t;\alpha_1,\beta_1')$.

**B.3':** $\pi_{ASA1} \equiv P(PP(1)=1, T(0)>\tau_0)$ is fixed and known.

Again we model each $w_z(\cdot)$ as in (4). The sensitivity parameters $\gamma = (\beta_0', \beta_1', \pi_{ASA1})$ have different interpretations than those under assumption set A; for example $\beta_0'$ is the log odds ratio of meeting the PP-criteria under $Z=1$ per unit increment of the failure time for subjects with $T>\tau_0$ under $Z=0$. The parameter $\pi_{ASA1}$ is constrained between $\max\{0, S_0(\tau_0)+pp_1-1\}$ and $\min\{pp_1, S_0(\tau_0)\}$. Algebra shows that

$$\int_{\tau_0}^{\infty} w_0(s;\alpha_0',\beta_0')dF_0^{\tau_0 obs}(s) = \pi_{ASA1}/S_0(\tau_0), \int_{\tau_0}^{\infty} w_1(s;\alpha_1',\beta_1')dF_1^{PPobs}(s) = \pi_{ASA1}/pp_1,$$
$$S_0^{ASA1}(t) = 1-(S_0(\tau_0)/\pi_{ASA1})\int_{\tau_0}^{t} w_0(s;\alpha_0',\beta_0')dF_0^{\tau_0 obs}(s),$$
$$S_1^{ASA1}(t) = 1-(pp_1/\pi_{ASA1})\int_{\tau_0}^{t} w_1(s;\alpha_1',\beta_1')dF_1^{PPobs}(s).$$

Nonparametric bounds for $SCE^{ASA1}(t)$ are constructed similarly as for $SCE^{APP}(t)$, replacing $F_0^{PPobs}(t)$ with $F_0^{\tau_0 obs}(t)$ and $pp_0$ with $S_0(\tau_0)$. With $\pi_{min} \equiv \max\{0, S_0(\tau_0)+pp_1-1\}$, the bounds for $S_z^{ASA1}(t)$ are informative if $\pi_{min}>0$, and for $t \geq \tau_0$ equal

$$S_{0,lo}^{ASA1}(t) \equiv \max\left\{0, 1-\frac{F_0^{\tau_0 obs}(t)}{\pi_{min}/S_0(\tau_0)}\right\}; S_{0,hi}^{ASA1}(t) \equiv \min\left\{\frac{S_0(\tau_0)}{\pi_{min}} - \frac{F_0^{\tau_0 obs}(t)}{\pi_{min}/S_0(\tau_0)}, 1\right\},$$

$$S_{1,lo}^{ASA1}(t) \equiv \max\left\{0, 1 - \frac{F_1^{\tau_0 obs}(t)}{\pi_{min}/pp_1}\right\}; S_{1,hi}^{ASA1}(t) \equiv \min\left\{\frac{pp_1}{\pi_{min}} - \frac{F_1^{\tau_0 obs}(t)}{\pi_{min}/pp_1}, 1\right\}.$$

*Assumption sets B, C, D.* Under AM included in B, C, and D, $S_z^{APP}(t) = S_z^{ASA1}(t)$, such that the identical methods described above for the $APP$ estimand apply.

**4.3. Estimands $S_1^{PP1}(t)$ and $S_0^{PP1}(t)$.**

The base set of assumptions identify $S_1^{PP1}(t)$ (by $S_1^{PPobs}(t)$), but not $S_0^{PP1}(t)$.

*Assumption set A.* We write

$$S_0^{PP1}(t) = P(PP(1) = 1|T(0) > t)S_0(t)/pp_1, \tag{11}$$

and use a selection bias model parallel to B.1 and B.1':

**B.1":** $P(PP(1) = 1|T(0) = t) = w_0(t; \alpha_0'', \beta_0'')$,

where again we specify $w_0(\cdot)$ as modified inverse-logit as in (4).

Straightforward calculation using (11) and B.1" shows that

$$S_0^{PP1}(t) = 1 - \frac{1}{pp_1}\int_0^t w_0(s; \alpha_0'', \beta_0'')dF_0(s), \tag{12}$$

where $\alpha_0''$ is the solution to $\int_0^\infty w_0(s; \alpha_0'', \beta_0'')dF_0(s) = pp_1$.

Since $S_1^{PP1}(t)$ is identifed from the base set of assumptions, nonparametric bounds for $SCE^{PP1}(t)$ derive from nonparametric bounds for $S_0^{PP1}(t)$. By similar derivations used in the proof of (8) in the Supplementary Materials, the bounds equal

$$\max\left\{0, 1 - \frac{[1 - S_0(t)]}{pp_1}\right\} \leq S_0^{PP1}(t) \leq \min\left\{\frac{S_0(t)}{pp_1}, 1\right\} \quad \text{for} \quad t \geq \tau_0. \tag{13}$$

The width of the bounds increases with the risk of early failure or nonadherence. In many applications, $S_0(t)/pp_1 > 1$ for most or all $t$, such that the upper bound is one.

*Assumption set B (add AM).* Under AM, $S_0^{PP1}(t) = P(T(0) > t|PP(1) = 1, H(0) = 1)$ for $t \geq \tau_0$, where $\{H(0) = 1\} = \{T(0) \leq \tau_0\} \cup \{PP(0) = 1\}$. With $F_0^{H(0)=1}(\cdot)$ the distribution of $T(0)$ conditional on $H(0) = 1$, we use the selection model:

**B.2":** $P(PP(1) = 1|H(0) = 1, T(0) = t) = w_0(t; \alpha_0'', \beta_0'')$ for $t \geq 0$, such that

$$S_0^{PP1}(t) = 1 - \frac{P(H(0) = 1)}{pp_1}\int_0^t w_0(s; \alpha_0'', \beta_0'')dF_0^{H(0)=1}(s) \quad \text{for} \quad t \geq \tau_0. \tag{14}$$

Simple calculations show that $P(H(0) = 1) = 1 - S_0(\tau_0) + pp_0$ and $F_0^{H(0)=1}(t) = \left(1 - S_0(\tau_0) + pp_0 F_0^{PPobs}(t)\right) / \left(1 - S_0(\tau_0) + pp_0\right)$ for $t \geq \tau_0$. With $w_0(\cdot)$ again the inverse-logit function, $\beta_0''$ is the log odds of $PP(1) = 1$ given $H(0) = 1$ and $T(0) = t$ versus $t - 1$. In many applications $\{H(0) = 1\} \approx \{PP(0) = 1\}$ due to a much larger number of subjects in $\{PP(0) = 1\}$ than in $\{T(0) \leq \tau_0\}$, in which case $\beta_0''$ has an interpretation very close to that of $\beta_0$ for the $APP$ estimand under assumption set A. (This close approximation attains in the Thai trial, where $\widehat{pp}_0 = 0.777$ and $1 - \widehat{S}_0(\tau_0) = 0.0012$.)

The nonparametric bounds are $S_{0,lo}^{PP1}(t) \leq S_0^{PP1}(t) \leq S_{0,hi}^{PP1}(t)$, where

$$\begin{aligned} S_{0,lo}^{PP1}(t) &= \max\left\{0, 1 - \left(1 - S_0(\tau_0) + pp_0 F_0^{PPobs}(t)\right) / pp_1\right\}, \\ S_{0,hi}^{PP1}(t) &= \min\left\{pp_0 \left(1 - F_0^{PPobs}(t)\right) / pp_1, 1\right\} \qquad \text{for } t \geq \tau_0. \end{aligned} \tag{15}$$

The bounds (15) are at least as sharp as those computed under the base set, (13).

*Assumption set C.* When SM is added, the methods are identical to those under assumption set B. Under assumption sets B and C, $S_0^{PP1}(t) = S_0^{APP}(t) = S_0^{ASA1}(t)$ if $\pi_{APP} = pp_1$, which occurs if $PP(1) = 1$ implies $PP(0) = 1$. Moreover, $S_1^{PP1}(t) = S_1^{APP}(t) = S_1^{ASA1}(t)$ if $\pi_{APP} = pp_1$ or $PP(0) \perp T(1) | PP(1) = 1$ (i.e., $\beta_1 = 0$ in B.2).

*Assumption set D (add SM, EA).* Under AM, $S_0^{PP1}(t) = S_0^{APP}(t)\pi_{APP}/pp_1$. SM and EA imply $\pi_{APP} = pp_0$ and $S_0^{APP}(t) = S_0^{PPobs}(t)$, such that $S_0^{PP1}(t)$ is identified by

$$S_0^{PP1}(t) = S_0^{PPobs}(t)\,(pp_0/pp_1) \qquad \text{for } t \geq \tau_0, \tag{16}$$

and there are no sensitivity parameters. Therefore under SM and EA, the non-causal estimand $SCE^{PPobs}(t) \equiv h(S_1^{PPobs}(t), S_0^{PPobs}(t))$ is corrected to a causal one through the adjustment constant $pp_0/pp_1$, which is $\leq 1$. The formula shows how the correction depends on differential rates of early failure and adherence. Under assumption set D, $S_0^{APP}(t) = S_0^{ASA1}(t) = S_0^{PP1}(t) = S_0^{PPobs}(t)$ if and only if $pp_0 = pp_1$, which is testable.

Table 2 lists the four assumption sets for each of the three estimands.

Table 2: Assumption Sets Under Which the Sensitivity Analysis May Be Performed

| Assumption Sets: | S. Pars. $\gamma$ | $\Gamma$ (Maximum Possible– Nonparametric Bounds) | $\Gamma$ (Maximum Plausible*) |
|---|---|---|---|
| | | **Estimand** $SCE^{APP}(t;\gamma)$ | |
| A. B.1, B.2, B.3 | $\beta_0$, $\beta_1$ | $-\infty < \beta_z < \infty$ | $\frac{1}{B} \leq \exp(\beta_z \bar{t}) \leq B \quad z = 0, 1$ |
| | $\pi_{APP}$ | $\max\{0, pp_0 + pp_1 - 1\} \leq \pi_{APP} \leq \min\{pp_0, pp_1\}$ | $pp_1 pp_0 \leq \pi_{APP} \leq \min\{pp_0, pp_1\}$ |
| B. B.1, B.2, B.3 | $\beta_0$, $\beta_1$ | $-\infty < \beta_z < \infty$ | $\frac{1}{B} \leq \exp(\beta_z \bar{t}) \leq B \quad z = 0, 1$ |
| AM | $\pi_{APP}$ | $\max\{0, S_0(\tau_0) + pp_1 - 1\} \leq \pi_{APP} \leq \min\{pp_0, pp_1\}$ | same as maximum possible range |
| C. B.1, B.2, B.3, | $\beta_0$, $\beta_1$ | $-\infty < \beta_z < \infty$ | $\frac{1}{B} \leq \exp(\beta_z \bar{t}) \leq B \quad z = 0, 1$ |
| SM, AM | $\pi_{APP}$ | $\max\{0, S_0(\tau_0) + pp_1 - S_1(\tau_0)\} \leq \pi_{APP} \leq \min\{pp_0, pp_1\}$ | same as maximum possible range |
| D. B.2, SM, EA | $\beta_1$ | $-\infty < \beta_1 < \infty$ | $\frac{1}{B} \leq \exp(\beta_1 \bar{t}) \leq B$ |
| | | **Estimand** $SCE^{ASA1}(t;\gamma)$ | |
| A. B.1', B.2', B.3' | $\beta'_0$, $\beta'_1$ | $-\infty < \beta'_z < \infty$ | $\frac{1}{B} \leq \exp(\beta'_z \bar{t}) \leq B \quad z = 0, 1$ |
| | $\pi_{ASA1}$ | $\max\{0, S_0(\tau_0) + pp_1 - 1\} \leq \pi_{ASA1} \leq \min\{S_0(\tau_0), pp_1\}$ | $pp_1 S_0(\tau_0) \leq \pi_{ASA1} \leq \min\{S_0(\tau_0), pp_1\}$ |
| B. B.1', B.2', B.3', | $\beta'_0$, $\beta'_1$ | $-\infty < \beta'_z < \infty$ | $\frac{1}{B} \leq \exp(\beta'_z \bar{t}) \leq B \quad z = 0, 1$ |
| AM | $\pi_{ASA1}$ | same as in B. for $APP$ | same as in B. for $APP$ |
| C. B.1', B.2', B.3', | $\beta'_0$, $\beta'_1$ | $-\infty < \beta'_z < \infty$ | $\frac{1}{B} \leq \exp(\beta'_z \bar{t}) \leq B \quad z = 0, 1$ |
| SM, AM | $\pi_{ASA1}$ | same as in C. for $APP$ | same as in C. for $APP$ |
| D. B.2', SM, EA | $\beta'_1$ | $-\infty < \beta'_1 < \infty$ | $\frac{1}{B} \leq \exp(\beta'_1 \bar{t}) \leq B$ |
| | | **Estimand** $SCE^{PP1}(t;\gamma)$ | |
| A. B.1" | $\beta''_0$ | $-\infty < \beta''_0 < \infty$ | $\frac{1}{B} \leq \exp(\beta''_0 \bar{t}) \leq B$ |
| B. B.2", AM | $\beta''_0$ | $-\infty < \beta''_0 < \infty$ | $\frac{1}{B} \leq \exp(\beta''_0 \bar{t}) \leq B$ |
| C. B.2", SM, AM | $\beta''_0$ | $-\infty < \beta''_0 < \infty$ | $\frac{1}{B} \leq \exp(\beta''_0 \bar{t}) \leq B$ |
| D. SM, EA | None | N/A (Full Identifiability) | N/A |

*$B > 1$ is a specified constant; $\bar{t}$ is a specified interpretable time-increment such as 1 year or the fixed follow-up period.

## 5. Maximum Likelihood Estimation and Uncertainty Intervals

We consider inference about the estimands $SCE^{\#}(t)$. Estimators of $SCE^{\#}(t;\gamma)$ for a fixed value of the sensitivity parameter(s) $\gamma$ are described in Section 5.1. Because $\gamma$ is not identifiable, we recommend reporting an estimated ignorance interval, i.e., a set of point estimates $\widehat{SCE}^{\#}(t;\gamma)$ with $\gamma$ varied over a selected region $\Gamma$. Ignorance intervals express ambiguity about the parameter of interest due to partial identifiability; in contrast, traditional confidence intervals (given a particular fixed $\gamma$) express ignorance/ambiguity due to sampling variability only. Associated with ignorance regions, in Section 5.2 we describe procedures for constructing uncertainty intervals that incorporate imprecision due to sampling variability as well as to lack of identifiability.

### 5.1. Estimation of $SCE^{APP}(t)$, $SCE^{ASA1}(t)$, and $SCE^{PP1}(t)$.

Given fixed $\gamma$, for each estimand and assumption set, maximum likelihood estimators (MLEs) of $SCE^{\#}(t;\gamma)$ are constructed by plugging in MLEs for the quantities given in Section 4. Kaplan-Meier estimates (nonparametric MLEs) can be used to estimate $F_z^{PPobs}(t)$, $F_z^{\tau_0 obs}(t)$, and $F_z(t)$, for $z = 0, 1$. MLEs for $pp_z$ are the proportion of subjects assigned treatment $z$ with $PP(z) = 1$. Estimation proceeds by fixing sensitivity parameters and then solving for unknown parameters. For example, for the $APP$ estimand under assumption set A, using (6) and fixing $\beta_z$ and $\pi_{APP}$, the MLE of $\alpha_z$ is the solution to $\int_{\tau_0}^{\infty} w_z(s;\alpha_z,\beta_z) d\widehat{F}_z^{PPobs}(s) = \pi_{APP}/\widehat{pp}_z$. Then, the MLE of $S_z^{APP}(t;\gamma)$ is obtained by plugging the MLEs into (7). The MLEs of the nonparametric bounds are obtained by substituting MLEs into (8) and (9). Estimation of the other estimands and under different assumption sets proceeds in a similar fashion.

As shown in Table 2, the sensitivity parameters $\pi_{\#}$ are constrained by functions of the identifiable parameters $pp_1$, $pp_0$, and $S_0(\tau_0)$. Therefore the data may contradict certain choices of $\pi_{\#}$, precluding estimation via plug-in MLEs in the above equations. Accordingly, we suggest using the MLEs of $pp_1$, $pp_0$, and $S_0(\tau_0)$ in the inequality bounds, as listed in Supplementary Table 3. This approach is equivalent to re-parametrizing

with an unconstrained sensitivity parameter (Supplementary Material, Section 3).

The data may violate an assumption that is nonetheless assumed. For example, assumption set D implies $pp_0 \leq pp_1$, but in the Thai trial $\widehat{pp}_0 > \widehat{pp}_1$. In such cases the estimation procedure uses constrained MLEs; in this example, for the $PP1$ estimand, $\widehat{pp}_0/\widehat{pp}_1$ is forced to unity such that $\widehat{S}_0^{PP1}(t) = \widehat{S}_0^{PPobs}(t)$ by equation (16).

**5.2. Estimated Ignorance Intervals and Uncertainty Intervals.**

With $\Gamma$ a specified region of sensitivity parameter values $\gamma$ and $\gamma_0 \in \Gamma$ the true value, we consider inference and uncertainty interval estimation for $SCE^{\#}(t;\gamma_0)$ for fixed $t \geq \tau_0$. For each fixed $\gamma \in \Gamma$, the bootstrap can be used to estimate the variances of $\widehat{S}_z^{\#}(t;\gamma)$ and $\widehat{SCE}^{\#}(t;\gamma)$, as well as to construct percentile confidence intervals for $S_z^{\#}(t;\gamma)$ and $SCE^{\#}(t;\gamma)$. To compute estimated ignorance and uncertainty intervals, we use the fact that for each estimand and assumption set, $\widehat{SCE}^{\#}(t;\gamma)$ is monotone in $\gamma$. With one sensitivity parameter such as $\gamma = \beta_1$, monotonicity implies that the extreme estimates $\widehat{SCE}^{\#}(t;\gamma)$ are obtained by setting $\beta_1$ to its extreme values. With three sensitivity parameters such as $\gamma = (\beta_0, \beta_1, \pi_{APP})$, monotonicity implies that the extreme estimates $\widehat{SCE}^{\#}(t;\gamma)$ are obtained by setting each of the three parameters to an extreme value; for example, Section 4.1 described how to achieve this for the $APP$ estimand. For a given analysis [estimand, assumption set, contrast function $h(\cdot,\cdot)$], suppose the values $\gamma_l$ and $\gamma_u$ in $\Gamma$ yield the minimum and maximum estimates $\widehat{SCE}^{\#}(t;\gamma)$. The estimated ignorance interval is computed as $\left[\widehat{SCE}^{\#}(t;\gamma_l), \widehat{SCE}^{\#}(t;\gamma_u)\right]$. Supplementary Table 3 lists the data-dependent values of $\gamma_l$ and $\gamma_u$ to use in a standard plausible-range sensitivity analysis that we propose in Section 5.3.

Following Vansteelandt et al. (2006, Section 4.1), a $(1-\alpha)\times 100\%$ pointwise uncertainty interval is defined by random limits $L^{\#}(t)$ and $U^{\#}(t)$ satisfying

$$\inf_{\gamma\in\Gamma} P_{\mathcal{M}(\gamma)}(L^{\#}(t) \leq SCE^{\#}(t;\gamma) \leq U^{\#}(t)) \geq 1-\alpha,$$

where $P_{\mathcal{M}(\gamma)}$ indicates that probabilities are taken under $\mathcal{M}(\gamma)$. Such intervals are of

interest because they contain the true $SCE^{\#}(t;\gamma_0)$ with probability at least $1-\alpha$. Imbens and Manski (2004) and Vansteelandt et al. (2006, Section 4.1) provided a method for obtaining estimated pointwise uncertainty intervals (henceforth EUIs), and showed that the EUIs have asymptotically correct coverage under the assumptions that $(\widehat{SCE}^{\#}(t;\gamma_l), \widehat{SCE}^{\#}(t;\gamma_u))^T$ is asymptotically normal and $\gamma_l$ and $\gamma_u$ are independent of the observed data law. If $\gamma_l$ and $\gamma_u$ are selected based on the observed data (as done, for example, with our standard plausible-range sensitivity analysis proposed below), then the bootstrap can be used to provide approximately correct EUIs (Vansteelandt et al. 2006, page 971); the bootstrap accounts for the sampling variability in $\gamma_l$ and $\gamma_u$ by recomputing $\gamma_l$ and $\gamma_u$ within each bootstrap iteration (see Section 3 in the Supplementary Material). Specifically, a bootstrap estimated $(1-\alpha)\times 100\%$ EUI for $SCE^{\#}(t;\gamma_0)$ is constructed as the union of the two bootstrap percentile $(1-\alpha)\times 100\%$ 1-sided confidence intervals for $SCE^{\#}(t;\gamma)$ computed at $\gamma=\gamma_l$ and at $\gamma=\gamma_u$, respectively. Here the significance level $\alpha$ does not need to be divided by two provided $SCE^{\#}(t;\gamma_u) > SCE^{\#}(t;\gamma_l)$ (Imbens and Manski, 2004), which occurs for all of the proposed sensitivity analysis methods because $\gamma_l \neq \gamma_u$ and $SCE^{\#}(t;\gamma)$ is strictly monotone in $\gamma$. Based on similar derivations used in Shepherd, Gilbert, and Lumley (2007), asymptotic normality of $(\widehat{SCE}^{\#}(t;\gamma_l), \widehat{SCE}^{\#}(t;\gamma_u))^T$ holds under simple conditions (provided in Supplementary Table 4), and their simulation studies plus those of SGD demonstrate satisfactory performance of the bootstrap, including nominal coverage probabilities of confidence intervals given fixed values of $\gamma$.

The null hypothesis $H_0^{inf}: \inf_{\gamma\in\Gamma}|SCE^{\#}(t;\gamma) - SCE^{*}| = 0$ for some fixed constant $SCE^{*}$ can be tested based on the EUI, by rejecting $H_0^{inf}$ if the $(1-\alpha)\times 100\%$ EUI excludes $SCE^{*}$. The corresponding 2-sided p-value equals the smallest $\alpha$ such that the $(1-\alpha)\times 100\%$ EUI excludes $SCE^{*}$. This testing procedure has size bounded above by $\alpha$, and thus has the desirable 1:1 correspondence with the EUIs. Alternatively, $H_0^{inf}$ can be tested directly using the infimum test of Todem et al. (2010), and their test

inverted to obtain an $(1-\alpha)\times 100\%$ EUI for $SCE^{\#}(t,\gamma_0)$, namely the set of $SCE^{*}$ where $H_0^{inf}$ is not rejected at significance level $\alpha$.

### 5.3. A Suggested Standard Sensitivity Analysis.

Given that it is often challenging to obtain consensus among experts about the plausible range of sensitivity parameters, it may be useful to develop standards that, while somewhat arbitrary, can be uniformly applied and make causal per-protocol analyses more objective and interpretable. We suggest one possible standard, where the odds ratios $\exp(\beta_0)$ and $\exp(\beta_1)$ are varied between $1/B$ and $B$ per increment $\bar{t}$ in failure time $T(0)$ and $T(1)$, where the scalar $\bar{t}$ and the bound $B$ are selected by subject matter experts based on the context of the study (following the approach of Shepherd, Gilbert, and Mehrotra, 2007). For example, $\bar{t}$ may simply be a 1-year interval. The plausible upper bounds for $\pi_{APP}$ and $\pi_{ASA1}$ are set at $\min\{\widehat{pp}_0, \widehat{pp}_1\}$ and $\min\{\widehat{S}_0(\tau_0), \widehat{pp}_1\}$, respectively, the MLEs of the maximum possible value. For assumption set A we suggest using the more plausible lower bound for $\pi_{APP}$ of $\widehat{pp}_1\widehat{pp}_0$, corresponding to the MLE of $\pi_{APP}$ if $PP(0)$ and $PP(1)$ were independent (negative correlation seems highly implausible). Similarly, we suggest using $\widehat{pp}_1\widehat{S_0}(\tau_0)$ as the plausible lower bound for $\pi_{ASA1}$. Under assumption sets B and C we suggest setting $\pi_{APP}$ and $\pi_{ASA1}$ at the MLEs of the minimum possible values, which vary by assumption set, since AM and SM greatly constrain the ranges for $\pi_{APP}$ and $\pi_{ASA1}$ (values listed in Supplementary Table 3). We also suggest reporting the nonparametric bounds as worst-case scenarios.

## 6. Example

We apply the above methodology to evaluate per-protocol vaccine efficacy in the Thai trial, using the vaccine efficacy contrast function $h(x,y) = 1-(1-x)/(1-y)$. The protocol definition of adherence required receiving all six immunizations at the four scheduled immunization visits within allowable visit windows by $\tau_0 \equiv 6.21$ months after randomization. The trial followed subjects for HIV infection up to the

terminal study visit at 42 months, and we focus on a time-point near the end of evaluation, $t = 39$ months. Thus, we denote the estimands of interest as $VE^{\#}(39;\gamma_0) = SCE^{\#}(39;\gamma_0)$. First we illustrate application of all of the methods, and second we evaluate the plausibility of each extra assumption, and, based on this assessment, conduct a substantively relevant sensitivity analysis. We use the yearly scale $\bar{t} = 1$ to define the maximum plausible selection bias odds ratio of $1/B$ or $B$, with $B = 1.5$. We construct EUIs and tests for $H_0^{inf}$ using the approach described in Section 5.2.

The choice of $B$ is important as it strongly affects the width of ignorance intervals and EUIs. Given the authors experience in HIV vaccine efficacy trials, and our lack of involvement in RV144 until after the primary publication, we counted ourselves as suitable experts to specify $B$. Because the infection rates by $\tau_0$ were very low, per-protocol status is approximately equivalent to adherence status, such that $B$ essentially measures the maximal association, within individuals adherent under the assigned treatment arm, of their infection times with adherence under the opposite treatment arm. Our choice of $B = 1.5$ reflects our belief that infection time had low association with adherence under the un-assigned arm in subjects already known to be adherent under the assigned arm. Moreover, our choice of relatively small $B$ was influenced by the fact that infection times are the outcome of many constituent factors including demographics, host genetics, sexual behavior, and prevalence/characteristics of HIV in sexual partners, and these factors had limited heterogeneity in this general population study with low representation of extremely high risk groups and with low observed heterogeneity in infection risk (Rerks-Ngarm et al. 2009). Given the challenge in selecting $B$, we recommend reporting results for both the plausible range and nonparametric bounds sensitivity analyses.

Figure 1 shows estimated ignorance intervals and 95% EUIs for $VE^{\#}(39;\gamma_0)$ for the three estimands under each assumption set. The methods are applied using both the maximum possible sensitivity parameter ranges (nonparametric bounds) and the max-

imum plausible ranges according to the suggested standard (semiparametric sensitivity analysis). For the nonparametric bounds approach, for each estimand and under assumption sets A, B, and C, the range of point estimates is large, extending to negative infinity. However, under assumption set D the estimated ignorance interval collapses to a single point, the point estimate under the naive analysis that uses the estimator $h(\widehat{S}_1^{PPobs}(39), \widehat{S}_0^{PPobs}(39))$. The plausible-range results show that the AM assumption (but not SM) leads to substantially narrower estimated ignorance intervals and EUIs. In addition, under assumption sets A, B, and C, these intervals are much narrower for the plausible-range sensitivity analysis than for the nonparametric bounds analysis, showing a major benefit from leveraging plausible assumptions.

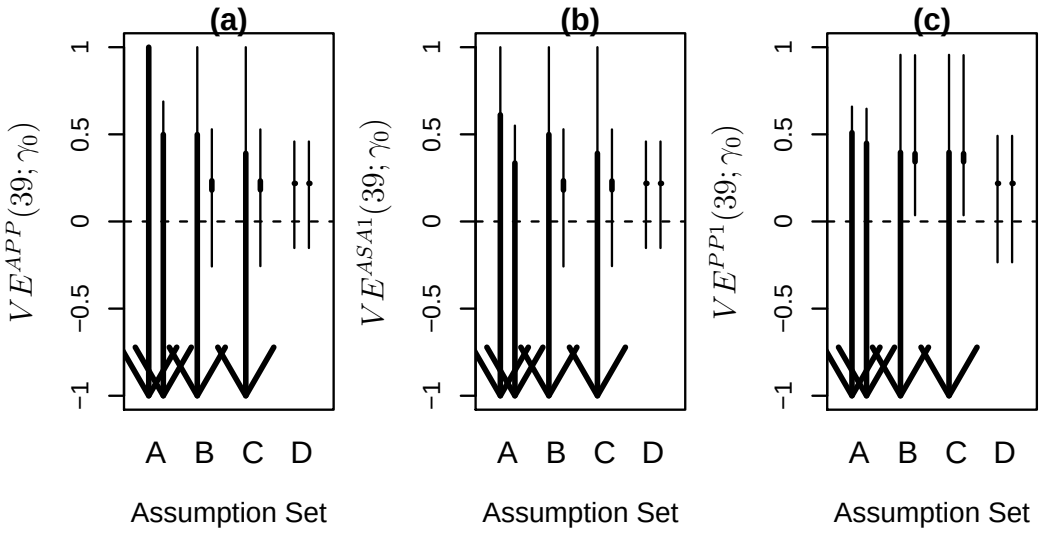


Figure 1. Estimated ignorance intervals (thick segments) and 95% EUIs (thin segments) for (a) $VE^{APP}(39;\gamma_0)$, (b) $VE^{ASA1}(39;\gamma_0)$, (c) $VE^{PP1}(39;\gamma_0)$ under assumption set A (base set), B (base set + AM), C (base set + SM, AM), D (base set + SM, EA). Left (right) vertical lines are nonparametric bounds (semiparametric bounds with plausible ranges), where arrows indicate extension to $-\infty$ (to a negative number $<$ -1.0).

Using assumption set A, Figure 2 shows the point and 95% confidence interval estimates for $VE^{APP}(39;\gamma)$ as a function of the fixed sensitivity parameters $\gamma =$

$(\beta_1, \beta_0, \phi_{APP})$, varied over the suggested standard plausible region $\Gamma$. As indicated by the dark shaded region, for the analysis to demonstrate significant benefit $VE^{APP}(39;\gamma) > 0$, the sensitivity parameters must satisfy $\phi_{APP} < 0.95$ and $(\exp(\beta_1), \exp(\beta_0))$ in the upper-left triangle with large values of $\exp(\beta_1)$ and small values of $\exp(\beta_0)$. Using estimates $\widehat{pp}_1$, $\widehat{S}_0(\tau_0)$, and $\widehat{S}_1(\tau_0)$, and under assumption sets B and C, $\phi_{APP}$ is constrained to be within $[0.9984, 1]$ and $[0.9992, 1]$, respectively. Hence, a sensitivity analysis under B or C corresponds approximately to Figure 2, $\phi_{APP} = 1.0$. The reduced variability of the vaccine efficacy estimate over the range induced by AM is apparent.

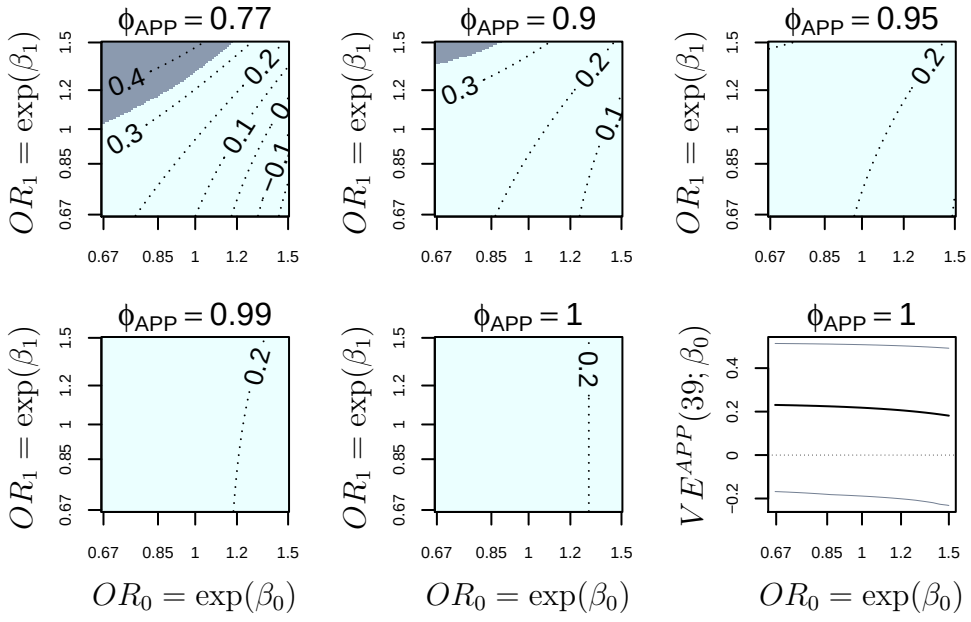


Figure 2. Contour plots of $\widehat{VE}^{APP}(39;\gamma)$ for $\gamma = (\beta_1, \beta_0, \phi_{APP})$ varying over the plausible region $\Gamma$, under assumption set A. Dark shaded regions indicate $\gamma$ values at which $VE^{APP}(39;\gamma)$ is significantly $> 0$. $\phi_{APP} = 0.77$ corresponds to $PP(1) \perp PP(0)$. When $\phi_{APP} = 1$, results only vary with $\beta_0$; thus the $OR_1$ parameter is irrelevant in the lower-middle contour plot, and the lower-right plot shows $\widehat{VE}^{APP}(39;\gamma)$ $(\gamma = \beta_0)$ with 95% confidence intervals as a function of $\exp(\beta_0)$.

Figure 3 shows the parallel results for $VE^{ASA1}(39;\gamma)$, indicating significant benefit

$VE^{ASA1}(39;\gamma) > 0$ for values of $\exp(\beta_0')$ less than about 0.7. The odds ratio $\exp(\beta_0')$ being less than 1 reflects the premise that placebo recipients with longer infection times have lower odds of being per-protocol under vaccine than placebo recipients with shorter infection times. This would be plausible if higher risk per-protocol subjects are less likely to be per-protocol if randomized to the other arm than lower risk per-protocol subjects. Thus if there is any evidence for a beneficial vaccine effect for the $ASA1$ subgroup, it is highly sensitive to small amounts of selection bias.

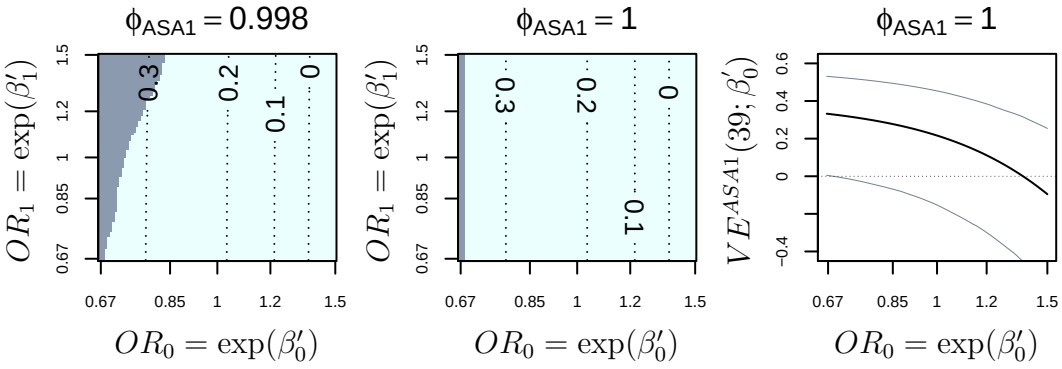


Figure 3. Contour plots (left and middle) of $\widehat{VE}^{ASA1}(39;\gamma)$ for $\gamma = (\beta_1', \beta_0', \phi_{ASA1})$ varying over the plausible region $\Gamma$, under assumption set A. Dark shaded regions indicate $\gamma$ values at which $VE^{ASA1}(39;\gamma)$ is significantly $> 0$. The right plot shows $\widehat{VE}^{ASA1}(39;\gamma)$ with 95% confidence intervals as a function of $\exp(\beta_0')$ when $\phi_{ASA1} = 1$.

Figure 4 shows the results for $VE^{PP1}(39;\gamma)$ under assumption set A and under B or C, demonstrating significant benefit $VE^{PP1}(39;\gamma) > 0$ for all values of $\exp(\gamma) = \exp(\beta_0'') <\approx 1.2$. Because $\beta_0''$ has a similar interpretation as $\beta_0$ for the $APP$ estimand, the significant effect is robust under a premise that per-protocol placebo recipients with shorter infection times would not have a higher odds of being per-protocol under vaccine than per-protocol placebo recipients with longer infection times.

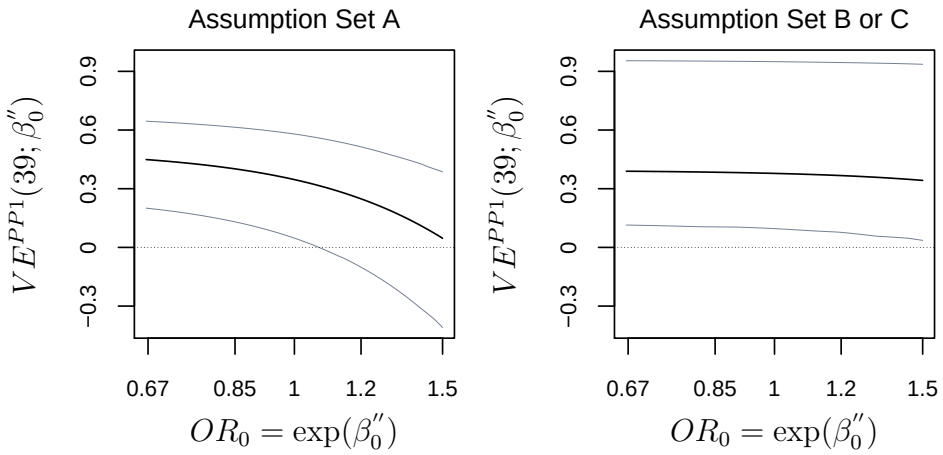


Figure 4. Estimates of $VE^{PP1}(39;\gamma)$ and 95% confidence intervals for $\gamma = \beta_0''$ values varying over the plausible range $\Gamma$, under assumption set A (base set; left panel) and under assumption set B or C (base set + AM or base set + SM, AM; right panel).

Next, we evaluate the plausibility of the identifiability assumptions. The consistency part of SUTVA holds because the trial is randomized and only one vaccine regimen was administered in a uniform manner, and the no interference part of SUTVA is plausible because the study sites were geographically dispersed across Thailand, the annual HIV incidence was low (0.2%), and only 1–2 clusters of infected subjects were identified. Ignorable treatment assignment is plausible because there were no recorded problems with the validity of the randomization. Random censoring is difficult to evaluate, but the impact of any violation of this assumption is minimized by the high rates of participant retention (10% drop-out by the terminal visit). SM is plausible, given that fewer subjects were observed to be infected by $\tau_0$ in the vaccine than placebo group (5 versus 10), with $\widehat{S}_1(\tau_0) = 0.9994$ and $\widehat{S}_0(\tau_0) = 0.9988$. AM is plausible but EA is not, given that, among those with an HIV negative test result at the Week 24 visit, the adherence rate was higher in the placebo than vaccine group: 5565 of 7317 (76.1%) versus 5285 of 7234 (73.1%) adherent (Chi-squared test $p < 0.0001$).

Assumption set C is therefore the strongest plausible assumption set, and we will

interpret the results based on this set. From Figure 1, the estimated ignorance interval (95% EUI) for per-protocol causal vaccine efficacy is 18% to 23% (-30% to 51%) for each of the $APP$ and $ASA1$ estimands, such that there is little evidence for positive vaccine efficacy after accounting for potential selection bias. In contrast, the estimated ignorance interval (95% EUI) for $VE^{PP1}(39;\gamma_0)$ is 34% to 39% (4% to 95%), indicating evidence for significant positive vaccine efficacy among those who were (or would have been) per-protocol when randomized to vaccine.

The hypothesis tests based on EUIs yield 2-sided p-values of 0.52 for the $APP$ and $ASA1$ estimands and 0.03 for the $PP1$ estimand.

To assess per-protocol vaccine efficacy over time, we repeated the above analyses for a grid of fixed times $t$ shortly after $\tau_0$ through to $\tau$ (9 to 39 months). Supplementary Figures 1–12 show the estimated ignorance intervals and 95% EUIs for $VE^{\#}(t;\gamma_0)$ and for the constituent survival curves $S_z^{\#}(t;\gamma_0)$, under each assumption set. Under sets B–D, the estimates of $VE^{\#}(t)$ decline over time for each estimand, with borderline significant $VE^{\#}(t) > 0$ through about 12–15 months for the plausible range analysis.

In conclusion, the causal sensitivity analysis provides a more interpretable assessment of per-protocol vaccine efficacy than the non-causal result originally reported in Rerks-Ngarm et al. (2009). Moreover, it provides a more complete account of uncertainty than the original analysis, by accounting for partial non-identifiability as well as for sampling variability. The original estimate of non-causal per-protocol vaccine efficacy was $\widehat{VE} = 25\%$, 95% CI -16% to 51%, p = 0.19, compared to the causal estimates of $\widehat{VE} = 18\%$ to 23% (for the $APP$ and $ASA1$ estimands), with 95% EUI -30% to 51%, p = 0.52. Thus, the causal analysis provides less evidence for beneficial vaccine efficacy through 39 months in subjects adherent to the full set of immunizations, in both the magnitude and precision of the point estimates. However, the time-dependent analysis indicates marginally significant positive per-protocol vaccine efficacy over the first 12–15 months after accounting for plausible levels of potential selection bias.

For the $APP$ estimand, the estimated ignorance intervals and EUIs are very wide under assumption set A, for both the semiparametric and nonparametric bounds, raising the question as to what are scenarios where the estimates will be sufficiently narrow to make precise conclusions. In general, both bounds will be narrow when the per-protocol probabilities and the event rates are high in both groups, in which case there is little room for selection bias. In the rare event setting such as for RV144, the semiparametric bounds may still be informatively precise if the per-protocol rates and numbers of events are large. To demonstrate this, Supplementary Figure 13 shows the identical analysis presented in Figure 1 for a second HIV vaccine efficacy trial with a very similar study design, Vax004 (Flynn et al. 2005), which had much higher per-protocol rates (92.6% and 93.0% in the vaccine and placebo groups) and 3-fold more events (368 HIV infections). The semiparametric estimated ignorance intervals and EUIs are much narrower than for RV144, demonstrating the impact of higher per-protocol rates and greater event numbers, respectively. However, the nonparametric bounds under assumption set A are still very wide, illustrating that in rare event settings the nonparametric bounds are typically uninformative.

## 7. Discussion

In clinical trials with survival time endpoint, the standard analysis of per-protocol treatment efficacy contrasts estimates of $S_1^{PPobs}(t)$ and $S_0^{PPobs}(t)$. Imbalances in predictors of the survival endpoint between the comparator groups $\{PP = 1, Z = 1\}$ and $\{PP = 1, Z = 0\}$ may easily occur, which renders the analysis non-causal and potentially misleading. Therefore, it is of interest to assess alternative per-protocol estimands that measure a causal effect of treatment. However, such estimands are not identifiable from the observable data plus standard assumptions, which makes a sensitivity analysis generally warranted, and motivates this work.

We defined three survival causal effect ($SCE$) per-protocol vaccine/treatment effi-

cacy parameters of interest– $SCE^{APP}(t)$, $SCE^{ASA1}(t)$, $SCE^{PP1}(t)$– and procedures for drawing inferences about these estimands under different assumption sets and specified regions for fixed sensitivity parameters. The estimands have different interpretations and hence different utilities. The $APP$ estimand may be particularly useful for guiding future research and policies for using the vaccine/treatment, given that always compliers (i.e., always per-protocol subjects) are those who can be expected to adhere to the treatment, and the level of efficacy in this subgroup is an important input parameter for models that predict the treatment's effectiveness in various settings where the adherence pattern may differ from that observed in the clinical trial. The $ASA1$ estimand measures the treatment effect for the always-survivors who would take treatment, regardless of their adherence under control, and may also be a useful input into treatment effectiveness models. Compared to the $APP$ estimand, the $ASA1$ estimand has advantage of identifiability under weaker assumptions, and disadvantage that the $ASA1$ subpopulation has a less straightforward interpretation, equal to the union of the always per-protocol subpopulation and the always-survivors who would adhere to $Z = 1$ but not to $Z = 0$. Reporting results for both estimands may be useful for assessing whether and to what extent treatment efficacy is associated with adherence under $Z = 0$. For blinded trials of non-toxic treatments, the estimands should be approximately equal, and in fact are exactly equal under the Adherence Monotonicity assumption that adherence to $Z = 1$ implies adherence to $Z = 0$. The $PP1$ estimand is least useful for evaluating treatment efficacy because by construction it favors the $Z = 1$ arm, such that its main value may be for assessing the amount of efficacy received for the observable subgroup of subjects who are per-protocol under active treatment $Z = 1$. An appeal of the $PP1$ estimand is that $Z = 1$ subjects in the $PP1$ cohort can be directly observed, and hence directly studied and characterized.

For trials with objective to evaluate per-protocol treatment efficacy, it is valuable to routinely report a sensitivity analysis of these causal estimands, at least for the

$APP$ and $ASA1$ estimands. Given that estimated uncertainty intervals are wider than confidence intervals, for trials where assessment of per-protocol treatment efficacy is a key objective, it may be prudent to increase the sample size accordingly, recognizing that increasing the sample size alone may not be sufficient to obtain narrow uncertainty intervals if the probability of being per-protocol is low.

Ideally, the per-protocol sensitivity analysis would account for potential bias due to unmeasured variables, after adjusting for potential bias due to observed variables. The sensitivity analysis methods developed in this article could be extended to adjust for observed covariates in a number of ways. One modification would be based on the fact that, for all assumption sets and for both the nonparameteric bounds and semiparametric sensitivity analysis approaches, the causal estimands $SCE^{\#}(t;\gamma_0)$ are functions of two types of terms: (1) fixed values of sensitivity parameters and models of selection bias; and (2) survival curves $S_z^{PPobs}(t)$, $S_z^{\tau_0 obs}(t)$, and $S_z(t)$ that are identified from the observed data under the base set of assumptions. Therefore, the inferential methods described in Section 5 may be modified to adjust for covariates by replacing the Kaplan-Meier estimates of the above survival curves with covariate-adjusted estimates, for example via the method described in Hernán (2010).

## Supplementary Materials

**Title: Supportive results** The supportive results include proofs, additional tables and figures describing the methods and results, and complete computer code.


## Acknowledgements

This research was supported by NIH NIAID grants 2 R37 AI054165-11 (PBG, MGH) and R01 DA023879 (BES). The authors thank the Editors and reviewers for very helpful comments, and the participants, investigators, and sponsors of the RV144 Thai trial, including the U.S. Military HIV Research Program (MHRP); U.S. Army Medical Research and Materiel Command; NIAID; U.S. and Thai Components, Armed

Forces Research Institute of Medical Science; Ministry of Public Health, Thailand; Mahidol University; SanofiPasteur; and Global Solutions for Infectious Diseases.

## References

Angrist, J., Imbens, G.W., and Rubin, D.B. (1996), "Identification of causal effects using instrumental variables (with discussion)," *Journal of the American Statistical Association*, 91, 444–472.

Baker, S. (1998),"Analysis of survival data from a randomized trial with all-or-none compliance: estimating the cost-effectiveness of a cancer screening program," *Journal of the American Statistical Association*, 93, 929–934.

Chiba, Y. (2012), "The large sample bounds on the principal strata effect with application to a prostate cancer prevention trial," *The International Journal of Biostatistics*, 8, Issue 1, Article 12.

Cuzick, J., Sasieni, P., Myles, J., and Tyler, J. (2007), "Estimating the effect of treatment in a proportional hazards model in the presence of non-compliance and contamination," *Journal of the Royal Statistical Society, Series B*, 69, 565–588.

Flynn, N.M., Forthal, D.N., Harro, C.D., et al. (2005), "Placebo-controlled trial of a recombinant glycoprotein 120 vaccine to prevent HIV infection," *Journal of Infectious Diseases*, 191, 654–665.

Follmann, D.A. (2000), "On the effect of treatment among would-be treatment compliers: an analysis of the multiple risk factor intervention trial," *Journal of the American Statistical Association*, 95, 1101–1109.

Frangakis, C.E., and Rubin, D.B. (2002), "Principal stratification in causal inference," *Biometrics*, 58, 22–29.

Gilbert, P.B., Berger, J.O., Stablein, D., Becker, S., Essex, M., Hammer, S.M., Kim, J., and DeGruttola, V.G. (2011), "Statistical interpretation of the RV144 HIV vaccine efficacy trial in Thailand: A case study for statistical issues in efficacy trials," *Journal of Infectious Diseases*, 203, 969–975.

Hernán, M.A. (2010), "The hazard of hazard ratios," *Epidemiology*, 21, 13–15.

Horne, A.D., Lachenbruch, P.A., and Goldenthal, K.L. (2001), "Intent-to-treat analysis and preventive vaccine efficacy," *Vaccine*, 19, 319–326.

Imbens, G.W., and Manski, C.F. (2004), "Confidence intervals for partially identified parameters," *Econometrica*, 72, 1845–1857.

Loeys, T., and Goetghebeur, E. (2003), "Causal proportional hazards estimator for the effect of treatment actually received in a randomized trial with all or nothing compliance," *Biometrics*, 59, 100–105.

Loeys, T., Goetghebeur, E., and Vandebosch, A. (2005), "Causal proportional hazards models and time-constant exposure in randomized clinical trials," *Lifetime Data Analysis*, 11, 435–449.

Nie., H., Cheng, J., and Small, D. (2011)."Inference for the effect of treatment on survival probability in randomized trials with noncompliance and administrative censoring," *Biometrics*, 67, 1397–1405.

Rerks-Ngarm, S., Pitisuttithum, P., Nitayaphan, S., et al. (2009), "Vaccination with ALVAC and AIDSVAX to prevent HIV-1 infection in Thailand," *New England Journal of Medicine*, 361, 2209–2220.

Robins, J.M. (1997), "Non-response models for the analysis of non-monotone non-ignorable missing data," *Statistics in Medicine*, 16, 21–37.

Rosenbaum, P.R. (1984), "The consequences of adjustment for a concomitant variable that has been affected by the treatment," *The Journal of the Royal Statistical Society, Series A*, 147, 656–666.

Scharfstein, D.O., Rotnitzky, A., and Robins, J.M. (1999), "Adjusting for nonignorable drop-out using semiparametric nonresponse models," *Journal of the American Statistical Association*, 94, 1096–1146.

Shepherd, B.E., Gilbert, P.B., and Dupont, C.T. (2011), "Sensitivity analyses comparing time-to-event outcomes only existing in a subset selected post-randomization and relaxing monotonicity," *Biometrics*, 67, 1100–1110.

Shepherd, B.E., Gilbert, P.B., and Lumley, T. (2007), "Sensitivity analyses comparing time-to-event outcomes only existing in a subset selected post-randomization," *Journal of the American Statistical Association*, 102, 573–582.

Todem, D., Fine, J., Peng, L. (2010), "A global sensitivity test for evaluating statistical hypotheses with nonidentifiable models," *Biometrics*, 66, 558–566.

Tsiatis, A.A., Davidian, M., Zhang, M., and Lu, X. (2008), "Covariate adjustment for two-sample treatment comparisons in randomized clinical trials: a principled yet flexible approach," *Statistics in Medicine*, 27, 4658–4677.

Vansteelandt, S., Goetghebeur, E., Kenward, M.G., and Molenberghs, G. (2006), "Ignorance and uncertainty regions as inferential tools in a sensitivity analysis," *Statistica Sinica*, 16, 953–979.

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: 353–368.