

Sensitivity Analyses Comparing Time-to-Event Outcomes Only Existing in a Subset Selected Post-Randomization, with Application to HIV Vaccine Trials

BRYAN E. SHEPHERD*, PETER B. GILBERT, THOMAS LUMLEY

*Department of Biostatistics,
University of Washington, Seattle, WA, 98195, USA*

*bshepher@u.washington.edu

SUMMARY

In preventative HIV vaccine efficacy trials it is of interest to determine whether randomization to vaccine affects post-infection outcomes that may be right censored. A few examples include time from infection diagnosis to initiation of antiretroviral therapy or time from infection diagnosis to AIDS. Here we present sensitivity analysis methods for making causal comparisons on these post-infection outcomes. We focus on estimating the survival causal effect, defined as the difference between probabilities of not yet experiencing the event in the vaccine and placebo arms, conditional on being infected regardless of treatment assignment. Our key assumption is monotonicity, that subjects randomized to the vaccine arm who become infected would have been infected if randomized to placebo. We propose non-parametric, semiparametric, and parametric methods which can be thought of as extending the work of Hudgens, Hoering, and Self (2003); Gilbert, Bosch, and Hudgens (2003); and Shepherd, Gilbert, Jemai, and Rotnitzky (2005) to handle right-censored outcomes. These methods are applied to the first Phase III preventative HIV vaccine trial (VaxGen's trial of AIDSVAX B/B).

Keywords: Causal inference; Principal stratification; Kaplan-Meier; AIDS.

1 INTRODUCTION

In preventative HIV vaccine efficacy trials it is of interest to assess vaccine effects on outcomes that occur only in infected individuals (Nabel, 2001; Graham, 2002; Gilbert *et al.*, 2005). Of such outcomes, some of the most important are time-to-event outcomes: for example the time from HIV infection diagnosis until the onset of AIDS or until the initiation of antiretroviral therapy (ART).

In efficacy trials, such as VaxGen's trial of AIDSVAX B/B, several thousands of individuals are randomized to either the vaccine or placebo arm. As in a typical clinical trial, they are followed for a certain amount of time, with the primary endpoint being HIV infection. Those that get infected are then enrolled in a post-infection study, and monitored for several additional years.

Comparisons of outcomes between vaccine and placebo recipients who become infected during the course of the trial could be misleading, because they condition on a post-randomization variable (infection) and therefore are susceptible to selection bias (Rosenbaum, 1984; Halloran and Struchiner, 1995). One could perform an intention-to-treat (ITT) analysis using all randomized individuals, ignoring infection status and defining the outcome of interest as the post-infection outcome (e.g., time from randomization until AIDS). However, this is problematic because the majority of trial participants will not get infected and are therefore not followed past the time that the initial study concludes, τ_0 . If one censors these uninfected individuals at τ_0 , then one induces dependent censoring and the Kaplan-Meier method provides a biased estimate that under-estimates the survival probability for $t > \tau_0$. On the other hand, if one assumes that the post-infection outcome occurs in none of the individuals who are not infected at time τ_0 and censors these individuals at the time when the post-infection study ends, $\tau_0 + \tau$, then one is most likely over-estimating the survival probability for $t > \tau_0$, particularly if τ is

large. An unbiased ITT analysis would be to censor everyone who does not experience the post infection event at τ_0 . This is, of course, unattractive because much information is thrown away. From a statistical standpoint, the ideal trial design would not have an initial stopping point, but would follow all individuals for the entire study time, $\tau_0 + \tau$. However, resource constraints make an efficacy trial of this type infeasible.

The method we propose here is based on principles of causal inference. Our analysis addresses a different question than an ITT analysis: among those who would have been infected regardless of treatment assignment, does assignment to the vaccine increase/decrease the probability of being AIDS free t months post-infection? It has been pointed out by many that comparing an outcome on the subgroup of individuals who would have been infected regardless of treatment assignment is a causal comparison (Kalbfleish and Prentice 1980; Robins 1995; Rubin 2000; Robins and Greenland 2000; Frangakis and Rubin 2002). This subgroup has been referred to as the always infected principal stratum (Hudgens, *et al.*, 2003; Gilbert, *et al.*, 2003). An analysis conditioning on membership in the always infected principal stratum is particularly of interest when one wants to know whether or not there exists a mechanism through which the vaccine alters a time-to-event outcome in infected individuals. Identifying and understanding any possible biological mechanisms may help future vaccine development. This also addresses a clinically important question: If a subject is going to be infected whether or not he takes the vaccine, will the vaccine be nonetheless beneficial? The current wave of candidate HIV vaccines undergoing efficacy testing have been specifically designed to alter disease progression/post-infection outcomes (www.hvtn.org).

Because one does not know which participants would have been infected regardless of treatment assignment, assumptions must be made to answer the causal question. Our key assumption is that any participant in the vaccine arm who becomes infected would have

been infected if randomized to placebo. Under this assumption, we advocate a sensitivity analysis approach where the probability of infection if assigned vaccine, given infection in the placebo arm and time from infection to some event of interest, is of a known form indexed by an assumed sensitivity parameter.

This work extends the sensitivity analysis methods reported in Hudgens, *et al.* (2003) (HHS); Gilbert *et al.* (2003) (GBH); and Shepherd, *et al.* (in press) (SGJR) to a time-to-event outcome defined post-randomization. In Section 2 we discuss notation, assumptions, and estimands; in Section 3 we discuss estimation; in Section 4 we explore the finite sample properties of our estimators; and in Section 5 we apply our methods to investigate the effect of vaccination on the time to ART initiation in VaxGen’s trial of AIDSVAX B/B. Section 6 discusses the methods and results. Technical details are found in the Appendix.

2 NOTATION, CAUSAL ESTIMAND, AND ASSUMPTIONS

Consider a study in which N subjects, independently and randomly selected from a given population of interest, are randomized to either placebo or vaccine. Let $Z_i = 1$ if subject i , $i = 1, \dots, N$, is randomized to vaccine and $Z_i = 0$ if randomized to placebo. Trial participants are monitored for HIV infection for a predetermined period of time; the infection status during the study follow-up period for subject i is the indicator S_i where $S_i = 1$ if infected and $S_i = 0$ if not. Let T_i be the time from infection diagnosis until some event for subject i , and define C_i as the time from infection diagnosis until censoring. We observe $Y_i = \min(T_i, C_i)$ and $\Delta_i = I(Y_i = T_i)$. Notice that T_i, C_i, Y_i , and Δ_i only exist if $S_i = 1$; otherwise they are assigned the value $*$.

To define the estimand of interest, we use potential outcomes/counterfactuals (Neyman 1923; Rubin 1978; Robins 1986). Specifically, define $S_i(0)$ to be the infection status indicator if, possibly contrary to fact, subject i is assigned placebo. Define $S_i(1)$ to be

the infection status indicator if subject i is assigned vaccine. Similarly, define $T_i(0)$ to be the time-to-event outcome if participant i is assigned placebo and $T_i(1)$ the time-to-event outcome if assigned vaccine. The potential outcomes $C_i(z)$, $Y_i(z)$, and $\Delta_i(z)$ are similarly defined for $z = 0, 1$. For a subject who does not become infected if assigned treatment z , i.e., $S_i(z) = 0$, we define $T_i(z) = Y_i(z) = C_i(z) = \Delta_i(z) = *$. This notation implicitly assumes that the potential outcomes of each trial participant are not influenced by the treatment assignments of other participants, an assumption known as the Stable Unit Treatment Value Assumption (SUTVA) (Rubin, 1978). SUTVA may be violated in vaccine studies of infectious disease when a local population is treated intensively, producing herd immunity. It is a reasonable assumption when study participants are geographically dispersed, as in our example.

Assuming the study participants make up a random sample from a large population of interest, the potential outcomes $W_i = (Z_i, S_i(0), S_i(1), T_i(0), T_i(1), C_i(0), C_i(1))$, $i = 1, \dots, N$, are i.i.d. copies of a random vector $W = (S(0), S(1), T(0), T(1), C(0), C(1), Z)$, and similarly the observed data $O_i = (Z_i, S_i, Y_i, \Delta_i)$, $i = 1, \dots, N$, are i.i.d. copies of $O = (Z, S, Y, \Delta)$. Randomization ensures that

$$(S(0), S(1), T(0), T(1), C(0), C(1)) \perp\!\!\!\perp Z \tag{2.1}$$

because $(S(0), S(1), T(0), T(1), C(0), C(1))$ can be considered an unobserved baseline characteristic of each subject. Here, for random variables A , B and C , $A \perp\!\!\!\perp B \mid C$ indicates conditional independence of A and B given C .

The four principal strata (Frangakis and Rubin, 2002) can be defined in terms of the counterfactual pair $(S(0), S(1))$: the never infected are those with $S(0) = S(1) = 0$, the harmed are those with $S(0) = 0$ and $S(1) = 1$, the protected are those with $S(0) = 1$ and $S(1) = 0$, and the always infected (*ai*) are those with $S(0) = S(1) = 1$.

For a subject i who is in the ai principal stratum, a causal effect on his/her time-to-event outcome is some measure of discrepancy between $T_i(0)$ and $T_i(1)$. For example, the difference $T_i(1) - T_i(0)$ was used by HHS and GBH, with the average causal effect in the ai stratum defined as $ACE = E[T(1) - T(0) | S(0) = S(1) = 1]$. With right-censored outcomes, such an estimand may not be identifiable without modeling the distribution of the time-to-event outcome; alternatively, one could compare restricted means (Chen and Tsiatis, 2001). However, it is often of primary interest to compare the probabilities of being event-free at time t , defined in the ai stratum as $1 - P(T_i(z) \leq t | S_i(0) = S_i(1) = 1)$ for $z = 0, 1$ (where $t = 0$ corresponds to the time of infection diagnosis). We take this approach; our estimand of interest is the “survival” causal effect in the ai stratum:

$$\begin{aligned} SCE(t) &= P(T_i(0) \leq t | S_i(0) = S_i(1) = 1) - P(T_i(1) \leq t | S_i(0) = S_i(1) = 1) \\ &\equiv F_p^{ai}(t) - F_v^{ai}(t). \end{aligned}$$

In order to estimate $SCE(t)$, we will make the following key assumptions:

A.1: Monotonicity: $S_i(1) \leq S_i(0)$.

A.2: Independent Censoring: $C_i(z) \perp\!\!\!\perp T_i(z) | S_i(z) = 1$.

The first of these assumptions implies that no individual is in the harmed principal stratum. This is a strong assumption but is quite plausible in randomized double-blinded placebo-controlled vaccine trials. The second assumption is common when analyzing time-to-event data, with the only difference being that this independence is made at the counterfactual level and is conditional on infection (otherwise $C_i(z) = T_i(z) = *$).

From these assumptions, $F_v^{ai}(t) = F_v(t)$, the distribution of the time-to-event outcome given vaccine and infection, and is therefore identified. However, in order to estimate $SCE(t)$, we must also be able to identify $F_p^{ai}(t)$ which is still not identifiable. The identifiable distribution, $F_p(t)$, is a mixture of the distribution of the time-to-event outcome for

placebos in the always infected stratum, $F_p^{ai}(t)$, and the distribution of the time-to-event outcome for placebos in the protected principal stratum, $F_p^{prot}(t)$. The mixing parameter is $VE \equiv 1 - P(S(1) = 1)/P(S(0) = 1)$, which under (2.1) and A.1 is one minus the probability of being in the always infected principal stratum.

From this information, one can write an expression for the bounds of $F_p^{ai}(t)$:

$$F_p^{ai,U}(t) = \min \left\{ \frac{F_p(t)}{1 - VE}, 1 \right\}, \quad (2.2)$$

$$F_p^{ai,L}(t) = \max \left\{ \frac{F_p(t) - VE}{1 - VE}, 0 \right\}. \quad (2.3)$$

This is equivalent to writing

$$F_p^{ai}(t) = (1 - VE)^{-1} \int_0^t w(s) dF_p(s),$$

where $w(t) \equiv P(S(1) = 1 | S(0) = 1, T(0) = t)$ and $w(t) = I_{\{t \leq q^{1-VE}\}}$ recovers $F_p^{ai}(t) = F_p^{ai,U}(t)$ and $w(t) = I_{\{t \geq q^{VE}\}}$ recovers $F_p^{ai}(t) = F_p^{ai,L}(t)$, with q^{1-VE} and q^{VE} being the $(1 - VE)^{th}$ and VE^{th} quantiles of $T(0)$, respectively.

There might be scientific reasons why these sharp bounds are too conservative. A different sensitivity analysis approach would be to use subject-matter knowledge to restrict the possible range for $F_p^{ai}(t)$. This was the approach of GBH and is specified by making the following additional assumption:

A.3: $P(S(1) = 1 | S(0) = 1, T(0)) = w(T(0); \beta)$, where $w(t; \beta) = \Phi \{ \alpha + g(t; \beta) \}$, β is fixed and known, $\Phi(\cdot)$ is a known cdf, α is an unknown parameter, and for each β , $g(\cdot; \beta)$ is a known function of T .

If one assumes A.3, then $F_p^{ai}(t)$ is identified. Of course, the parameter β is not identified by the observed data. It is regarded as fixed and known, and then varied over a plausible range of values as a form of sensitivity analysis (Scharfstein *et al.*, 1999; GBH). One choice for $w(t; \beta)$ is the expit function, i.e. $w(t; \beta) = (1 + \exp(-\alpha - \beta t))^{-1}$

used by GBH. For this choice of $w(\cdot)$, the sensitivity parameter β has a log odds ratio interpretation. Using the expit function and choosing $\beta = 0$ is the same as assuming $T(0) \mid S(1) \mid S(0) = 1$, or equivalently that the distribution of the time-to-event outcome under placebo is the same in the ai and in the protected principal strata. Choosing $\beta = -\infty$ and $\beta = \infty$ corresponds to the sharp bounds of HHS.

3 ESTIMATION

3.1 Non-Parametric Estimation: Sharp Bounds

Under SUTVA, (2.1), A.1, and A.2, and for $t < \tau$, where τ is the maximum post-randomization follow-up time, the sharp bounds given by (2.2) and (2.3) are consistently estimated as

$$\hat{F}_p^{ai,U}(t) = \min \left\{ \frac{\hat{F}_p(t)}{1 - \widehat{VE}}, 1 \right\},$$

$$\hat{F}_p^{ai,L}(t) = \max \left\{ \frac{\hat{F}_p(t) - \widehat{VE}}{1 - \widehat{VE}}, 0 \right\},$$

where $\widehat{VE} = \min\{1 - (n_v N_p)/(n_p N_v), 0\}$ ($N_z [n_z]$ as the number of individuals randomized to [infected in] treatment arm z , where $N_v + N_p = N$) and $\hat{F}_p(t)$ is the standard Kaplan-Meier estimator of $F_p(t)$. Since under monotonicity, $F_v^{ai}(t) = F_v(t)$, estimated sharp bounds for $SCE(t)$ are $\hat{F}_p^{ai,L}(t) - \hat{F}_v(t)$ and $\hat{F}_p^{ai,U}(t) - \hat{F}_v(t)$, where $\hat{F}_v(t)$ is the Kaplan-Meier estimator of $F_v(t)$.

Under certain conditions, $\hat{F}_p^{ai,U}(t)$ and $\hat{F}_p^{ai,L}(t)$ are asymptotically normal. Specifically, \widehat{VE} is asymptotically normal if $0 < VE < 1$, or equivalently $0 < P(S(0) = 1)$ and $0 < P(S(1) = 1) < P(S(0) = 1)$. The Kaplan-Meier estimate, $\hat{F}_p(t)$, is asymptotically normal under the usual conditions (Fleming and Harrington, 1991). Therefore, $\hat{F}_p^{ai,U}(t)$ and $\hat{F}_p^{ai,L}(t)$ are asymptotically normal if, in addition to these conditions,

$0 < F_p(t) < 1 - VE$ and $VE < F_p(t) < 1$, respectively. (These latter conditions are a result of (2.2) and (2.3). If these conditions are violated then, by definition, $F_p^{ai,U}(t)$ and $F_p^{ai,L}(t)$ are 1 and 0, respectively, and hence estimates will not be asymptotically normal.) This follows from the asymptotic normality of $(\hat{F}_p(\cdot), \widehat{VE})$, the Hadamard differentiability of the maps $\hat{F}_p(t)/(1 - \widehat{VE})$ and $(\hat{F}_p(t) - \widehat{VE})/(1 - \widehat{VE})$, and the functional delta method (Andersen *et al.*, 1992). Under these conditions, expressions for the asymptotic variance obtained via the functional delta method are

$$\begin{aligned} \text{var} \left(\hat{F}_p^{ai,U}(t) \right) &= \left(\frac{p_0}{p_1} \right)^2 \sigma^2(t) + \left(\frac{F_p(t)}{p_1} \right)^2 \frac{p_0(1-p_0)}{N_p} + \left(\frac{F_p(t)p_0}{p_1^2} \right)^2 \frac{p_1(1-p_1)}{N_v} \\ \text{var} \left(\hat{F}_p^{ai,L}(t) \right) &= \left(\frac{p_0}{p_1} \right)^2 \sigma^2(t) + \left(\frac{1-F_p(t)}{p_1} \right)^2 \frac{p_0(1-p_0)}{N_p} + \left(\frac{(1-F_p(t))p_0}{p_1^2} \right)^2 \frac{p_1(1-p_1)}{N_v}, \end{aligned}$$

where $\sigma^2(t)$ is the variance of the Kaplan-Meier estimate, $p_0 \equiv P(S = 1|Z = 0)$, and $p_1 \equiv p_0(1 - VE) = P(S = 1|Z = 1)$. From these equations one may estimate variances in the usual manner, by plugging in parameter estimates.

Variances may alternatively be estimated using a standard bootstrap procedure. Specifically, from (O_1, \dots, O_N) sample with replacement N vectors O_i , creating (O_1^*, \dots, O_N^*) . Compute $\widehat{SCE}^*(t)$ based on the bootstrap sample (O_1^*, \dots, O_N^*) . Repeat this process K times, and estimate the variance of $\widehat{SCE}(t)$ by the sample variance of the $\widehat{SCE}^*(t)$'s. $100(1 - \alpha)\%$ -level confidence intervals may be constructed using the resulting variance estimate (Wald-intervals), by the $\alpha/2$ - and $(1 - \alpha/2)$ -quantiles of $\widehat{SCE}^*(t)$ (percentile intervals), or by studentizing with the asymptotic variance estimate.

3.2 Parametric Estimation

In addition to SUTVA, (2.1), A.1, and A.2, assume A.3, and make the following modeling assumption:

- M.1: The distribution of $T(1)$ given $S(1) = 1$ is known up to a finite dimensional parameter η_1 ; that is, $f_{T(1)|S(1)=1}(t|S(1) = 1) = f_v(t; \eta_1)$ where η_1 is unknown and for each η_1 , $f_v(\cdot; \eta_1)$ is a known density.

Also make one of the following two assumptions on the distribution of $T(0)$:

- M.2a: The distribution of $T(0)$ given $S(0) = 1$, is known up to a finite dimensional parameter η_0^a ; that is, $f_{T(0)|S(0)=1}(t|S(0) = 1) = f_p(t; \eta_0^a)$, where η_0^a is unknown and for each η_0^a , $f_p(\cdot; \eta_0^a)$ is a known density.
- M.2b: The distribution of $T(1)$ given $S(0) = S(1) = 1$ is known up to a finite dimensional parameter η_0^b ; that is, $f_{T(0)|S(0)=S(1)=1}(t|S(0) = S(1) = 1) = f_p^{ai}(t; \eta_0^b)$, where η_0^b is unknown and for each η_0^b , $f_p^{ai}(\cdot; \eta_0^b)$ is a known density.

For ease of reference, we call the model defined by SUTVA, (2.1), A.1, A.2, A.3, M.1, and M.2a model \mathcal{M}_a . We call \mathcal{M}_b the model defined like \mathcal{M}_a except replacing M.2a with M.2b and demanding that $w(t; \beta) > 0$ for all t .

Under \mathcal{M}_a , $SCE(t)$ is a function of unknown parameters $(\alpha, \eta_0^a, \eta_1)$. Specifically, $SCE(t) = SCE_a(t; \alpha, \eta_0^a, \eta_1)$ where

$$SCE_a(t; \alpha, \eta_0^a, \eta_1) \equiv \frac{\int_0^t w(s; \alpha, \beta) f_p(s; \eta_0^a) ds}{\int_0^\infty w(s; \alpha, \beta) f_p(s; \eta_0^a) ds} - \int_0^t f_v(s; \eta_1) ds.$$

Similarly, under model \mathcal{M}_b , $SCE(t) = SCE_b(t; \eta_0^b, \eta_1) \equiv \int_0^t f_p^{ai}(s; \eta_0^b) ds - \int_0^t f_v(s; \eta_1) ds$.

The maximum likelihood estimators of $SCE(t)$ under \mathcal{M}_a and \mathcal{M}_b are therefore equal to the functions $SCE_a(t; \cdot, \cdot, \cdot)$ and $SCE_b(t; \cdot, \cdot)$ evaluated at the ML estimators of $(\alpha, \eta_0^a, \eta_1)$ and (η_0^b, η_1) under models \mathcal{M}_a and \mathcal{M}_b , respectively.

In the absence of censoring, the likelihood induced by these assumptions (minus A.2) is given in SGJR. This likelihood can be easily modified to account for independent post-

infection censoring:

$$\begin{aligned} \mathcal{L}_a(\alpha, \eta_0^a, \eta_1) &\propto \prod_{i=1}^N \left\{ \left[f_v(y_i; \eta_1)^{\delta_i} (1 - F_v(y_i; \eta_1))^{1-\delta_i} p_0 \int w(y; \alpha, \beta) f_p(y; \eta_0^a) dy \right]^{s_i} \right. \\ &\quad \times \left. \left[1 - p_0 \int w(y; \alpha, \beta) f_p(y; \eta_0^a) dy \right]^{1-s_i} \right\}^{z_i} \\ &\quad \times \left\{ \left[f_p(y_i; \eta_0^a)^{\delta_i} (1 - F_p(y_i; \eta_0^a))^{1-\delta_i} p_0 \right]^{s_i} (1 - p_0)^{1-s_i} \right\}^{1-z_i}, \quad (3.4) \end{aligned}$$

under model \mathcal{M}_a . Under model \mathcal{M}_b the likelihood $\mathcal{L}_b(\cdot)$ is defined as $\mathcal{L}_a(\cdot)$ but with $f_p^*(y; \alpha, \eta_0^b) \equiv \frac{w^{-1}(y; \alpha, \beta) f_p^{a_i}(y; \eta_0^b)}{\int w^{-1}(y; \alpha, \beta) f_p^{a_i}(y; \eta_0^b) dy}$ replacing $f_p(y; \eta_0^a)$.

Provided the protected principal stratum is non-empty, then under sufficiently smooth parameterizations the ML estimators of the model parameters are asymptotically normally distributed. The variance of the normal limiting distribution can be consistently estimated with either the observed or the (estimated) expected information. These, in turn, can be used in conjunction with the delta method to obtain consistent variance estimators of $SCE(t)$ for each fixed t . Sensitivity analyses are performed by varying the range of β .

It is worth noting that this likelihood can easily be extended to condition on baseline covariates. Likelihood based methods when the outcome of interest is not right censored were extensively studied in SGJR, and general principles stated there apply here.

3.3 Semi-Parametric Estimation

Consider estimation under SUTVA, (2.1), A.1, A.2, and A.3; i.e., performing sensitivity analyses by modeling $w(\cdot)$ but leaving the distributions of $T(0)$ and $T(1)$ unspecified. One can think of this as extending GBH to time-to-event outcomes.

The estimating equations GBH used to estimate $F_p^{ai}(t)$ when $n_v/N_v < n_p/N_p$ can be

written as

$$0 = \psi(p_0, \alpha) = \begin{cases} \sum_{i=1}^N (1 - Z_i)(S_i - p_0) \\ \sum_{i=1}^N Z_i \left(S_i - p_0 \int_0^\infty w(t; \alpha, \beta) d\hat{F}_p(t) \right). \end{cases} \quad (3.5)$$

Similar to Section 3.1, a natural approach would use these same equations, only now estimating $F_p(t)$ with the Kaplan-Meier estimate. In practice, however, this approach may not be feasible because $\hat{F}_p(t)$ is not well defined for $t > \tau$. This implies that the integral in the second estimating equation of (3.5) cannot be computed for $t > \tau$.

One way to fix this problem would be to assume some distributional form for the tail of $F_p(\cdot)$. This approach would be similar to the parametric methods of Section 3.2, however, and in this section we want to leave $F_p(\cdot)$ unspecified. Another approach would change the form of $w(\cdot)$, making it constant after time τ . For example, consider $w(\cdot)$ defined as follows:

$$w(t; \alpha, \beta) = \begin{cases} (1 + \exp(-\alpha - \beta t))^{-1} \text{ for } t \leq \tau \\ (1 + \exp(-\alpha - \beta \tau))^{-1} \text{ for } t > \tau. \end{cases} \quad (3.6)$$

Another choice for $w(\cdot)$ could be

$$w(t; \alpha, \beta) = (1 + \exp(-\alpha - \beta I_{\{t > t_0\}}))^{-1} \quad (3.7)$$

for some $t_0 \leq \tau$. Both (3.6) and (3.7) define $w(\cdot)$ with the expit function, but do so in a manner such that $w(\cdot)$ is constant for $t > \tau$, allowing one to write:

$$\int_0^\infty w(s; \alpha, \beta) d\hat{F}_p(s) = \int_0^\tau w(s; \alpha, \beta) d\hat{F}_p(s) + w(\tau; \alpha, \beta) (1 - \hat{F}_p(\tau)).$$

Of course, these choices of $w(\cdot)$ have implications with regards to interpretation. Under (3.6) the interpretation of β is technically the following: Given infection in the placebo arm, the odds of infection if randomized to the vaccine arm for $T = t_1$ versus

$T = t_2$ are $\exp\{\beta[\min(t_1, \tau) - \min(t_2, \tau)]\}$. This more complex interpretation might appear troublesome. However, over the range of t for which there are data, β has the usual log odds ratio interpretation. Under (3.7), β has a standard odds ratio interpretation, except now we have dichotomized $w(\cdot)$, assigning a particular probability of being in the ai stratum for $t > t_0$ and another for $t \leq t_0$.

With $w(\cdot)$ modeled by either (3.6) or (3.7), an extension of GBH using the Kaplan-Meier estimates for $F_p(t)$ is the semi-parametric MLE. In the appendix we show that under these same assumptions and $0 < VE < 1$, $p_0 > 0$, and for a properly specified well-behaved $w(\cdot)$ (i.e., constant for $t > \tau$, twice differentiable, and bounded), $\hat{F}_p^{ai}(t)$ is consistent and asymptotically normal for $t \in (0, \tau]$. Therefore, $\widehat{SCE}(t)$ is also consistent and asymptotically normal.

Because $\widehat{SCE}(t)$ is asymptotically normal, pointwise Wald-based confidence intervals based on the bootstrap will be valid for large sample sizes. It is also possible to obtain an analytic form for the asymptotic variance of $\widehat{SCE}(t)$. This variance estimate relies on being able to approximate $\hat{F}_p(t)$ with a sum of i.i.d. random variables. Such an approximation can be obtained from Stute (1995). Using this result, one can augment (3.5) by including additional estimating equations:

$$\{0, \dots, 0\}^T = \left\{ \sum_{i=1}^N (1 - Z_i) S_i (V_{1i} - F_p(t_1)), \dots, \sum_{i=1}^N (1 - Z_i) S_i (V_{ki} - F_p(t_k)) \right\}^T,$$

where k is the number of distinct failure times in the placebo arm, t_j is the j th ordered failure time, and for a specific j , V_{ji} are i.i.d. random variables for $i = 1, \dots, N$. One can then estimate the variance of parameter estimates using a sandwich estimator type approach, and from there one can estimate the variance of $\widehat{SCE}(t)$ using the delta method. Details are given in the appendix.

4 SIMULATIONS

To evaluate the small-sample performance of our estimators of $SCE(t)$ we conducted a 2×4 factorial simulation experiment, corresponding to generating data under $VE \equiv P(S(1) = 0 | S(0) = 1) \approx 0.3$ or 0.6 , and $\beta = 0.1, 0.2, 1$, or ∞ . Each simulation generated 1000 vectors W according to the following steps:

Step 1. The first 500 vectors were set at $Z = 0$, the second 500 were set at $Z = 1$.

Step 2. $S(0)$ was drawn from a Bernoulli(p_0) distribution with $p_0 = 0.25$ (this choice yields an expected number of infections in the placebo arm of 125, which is typical for a Phase III vaccine trial).

Step 3. $T(0)$ was generated for all realizations with $S(0) = 1$ according to the distribution $F_p(t; \eta)$ with $F_p(\cdot)$ a Weibull distribution and $\eta = (\text{shape} = 0.5, \text{scale} = 25)$. This distribution was chosen to reflect the distribution of the time from infection diagnosis to initiation of antiretroviral therapy (ART) in the VaxGen trial, in which approximately 50% of infected participants started ART by 24 months post-infection diagnosis.

Step 4. Given $T(0)$, for each realization with $Z = 1$ and $S(0) = 1$, $S(1)$ was drawn from a Bernoulli($w(T(0); \beta, \alpha)$) distribution. For $\beta = 0.1, 0.2$, and 1 , $w(t; \beta, \alpha)$ was defined as in (3.6) with $\tau = 24$ months. To ensure that $VE \approx 0.3$, α was set at $-0.2, -0.9$, or -3.6 , when β was set at $0.1, 0.2$, or 1 respectively; and to ensure that $VE \approx 0.6$, α was set at $-1.8, -3.4$, or -20 , when $\beta = 0.1, 0.2$, or 1 . For $\beta = \infty$, $w(t) = I_{\{t \geq q^{VE}\}}$ as discussed in Section 2, where $q^{0.3} = 3.18$ and $q^{0.6} = 21.0$.

Step 5. For the realizations with $Z = 1$ and $S(1) = 1$, $T(1)$ was set equal to $T(0)$.

Step 6. For the realizations with $S = 1$, C_1 was generated from a Weibull distribution with shape and scale parameters 3 and 35. Then C was set as $\min(\tau, C_1)$.

Step 7. For all realizations with $S = 1$, Y was chosen as $\min(C, T)$ and $\delta = I_{\{Y=T\}}$.

It is easily verified that these steps result in simulating data under SUTVA, (2.1), A.1, A.2, A.3, and model (3.6), with $SCE(t) = 0$ for all t .

For each simulated dataset, we computed $\hat{F}_p^{ai}(t)$ and $\widehat{SCE}(t)$ for $t = 24$ months, assuming the true model for $w(\cdot)$ and the true value for β . Wald-based 95% confidence intervals were constructed by estimating the standard error of estimates using both the bootstrap and asymptotic variance estimates.

Table 1 reports the performance of $\widehat{SCE}(t)$ based on 1000 simulation iterations. Since $\widehat{SCE}(t)$ is the difference between $\hat{F}_p^{ai}(t)$ and $\hat{F}_v(t)$, it is also useful to examine the performance of the estimates of $F_p^{ai}(t)$. Table 2 does this using the same simulations and analyses reported in Table 1. In addition to presenting the coverage of the untransformed 95% Wald-confidence intervals for $F_p^{ai}(t)$, Table 2 also presents confidence intervals by transforming symmetric confidence limits for $\log[-\log\{F_p^{ai}(t)\}]$.

In most cases, bias is minimal and coverage is good using either the bootstrap or the asymptotic variance estimate. The only exceptions are when $VE \approx 0.6$ and β is large. The poor coverage probabilities here are due to a boundary issue. Consider first the simulations with $\beta = \infty$. As discussed in Section 3.1, as $N \rightarrow \infty$, under the usual assumptions and if $VE < F_p(t)$ then $\hat{F}_p^{ai,L}(t)$ (which is equivalent to $\hat{F}_p^{ai}(t)$ with $\beta = \infty$) will be asymptotically normal and Wald-based confidence intervals will cover at their nominal level. However, in these simulations $F_p(t) = 0.625$ for $t = 24$ months, which is very close to $VE = 0.600$. Therefore, due to stochastic variation and our relatively small sample size, \widehat{VE} is often greater than $\hat{F}_p(t)$ resulting in $\hat{F}_p^{ai}(t) = 0$ in nearly half of the simulations. Consequently, the distribution of $\hat{F}_p^{ai}(t)$ is far from normal; therefore these confidence intervals that assume normality have poor coverage. (Notice that in this particular setting, Wald-based confidence intervals extended outside the $[0, 1]$ range; and log-log transformed confidence intervals could not be computed because the estimated

value of $F_p^{ai}(t)$ was often 0.) Interestingly, the 2.5- and 97.5- bootstrapped percentile confidence intervals for $F_p^{ai}(t)$ and $SCE(t)$ have coverage probabilities of 0.938 and 0.964, respectively, for the simulations with $VE \approx 0.6, \beta = \infty$.

The same logic explains why coverage and bias were also poor for the simulations with $VE \approx 0.6, \beta = 1$. Under these settings a value of $\beta = 1$ is quite large: for example, the odds of being infected in the vaccine arm given infection in the placebo arm from $t = 12$ and $t = 24$ (a difference of one year) multiplicatively increase $exp(12) \approx 163,000!$ Analyses based on the assumption that $\beta = 1$ are not too different from analyses assuming $\beta = \infty$. Again, the distribution of $\hat{F}_p^{ai}(t)$ is far from normal. Figure 1 shows a histogram of $\hat{F}_p^{ai}(t = 24)$ for the simulations with $VE \approx 0.6, \beta = 1$, as well as a similar histogram with $VE \approx 0.3, \beta = 1$ (where the method worked well) for purpose of comparison. Figure 1 also shows the true values of $F_p(t)$ and $F_p^{ai}(t)$ under these simulation settings. Notice how close $F_p^{ai}(t)$ is to 0 at $t = 24$ months.

5 EXAMPLE

We illustrate our methods using data from the VaxGen vaccine trial. This was a randomized, double-blind, placebo-controlled Phase III trial of AIDSVAX B/B conducted between 1998 and 2003. This study recruited 5,403 HIV negative, at risk individuals from 61 sites spanning large cities of North America and the Netherlands. The ratio of vaccine to placebo assignment was 2:1. Overall, the vaccine was not found to protect against HIV infection ($\widehat{VE} = 0.048$), although interaction tests suggested that the vaccine might partially prevent infection for non-whites. Among non-whites vaccine efficacy was estimated as 0.469. Detailed study results are found in Flynn *et al.* (2005). Here we compare the time from infection diagnosis to the initiation of antiretroviral therapy (ART) between the vaccine and placebo arms among participants (overall and within the non-white subgroup)

who would have been infected regardless of randomization assignment. Specifically, we perform sensitivity analyses to test the hypothesis $H_0 : SCE(t) = 0$ for $t = 1.17$ and 2 years, time points pre-specified for analysis by the VaxGen protocol. A vaccine effect to delay ART is beneficial to an individual because it delays exposure to drug toxicities, drug resistance, and the depletion of future therapy options.

A total of 368 subjects were infected during the trial, and of these, 347 enrolled into the post-infection study phase (225 in the vaccine arm). There was presumably little, if any, interaction between trial participants, so SUTVA seemed reasonable. As discussed in SGJR, A.1 was also thought to be plausible. In addition, the censoring mechanism did not appear to be informative based on similar drop-out rates for participants with different levels of behavioral risk; hence there is no evidence that A.2 is violated.

The three plots in Figure 2 show analyses looking at the time from infection diagnosis to the initiation of ART. Figure 2A shows the Kaplan-Meier estimates for both the vaccine and placebo arms for the probability of not yet starting ART. The plot also includes the estimates of the upper and lower bounds of $F_p^{ai}(t)$, described in Section 3.1. Figures 2B and 2C are semiparametric sensitivity analyses looking at $SCE(t)$ for $t = 1.17$ and 2 years. The plots contain both the estimate for $SCE(t)$ and 95% Wald confidence intervals (constructed using the asymptotic variance approximation and the bootstrap, with 500 bootstrap replications). In these analyses (and all other sensitivity analyses in this section), we modeled the probability of infection in the vaccine arm given infection in the placebo arm, $w(t; \alpha, \beta)$, with (3.6). Before performing these analyses, we elicited a plausible range for the sensitivity parameter β from a subject matter expert, Dr. Marc Gurwith of VaxGen. His best “guess” for a range for $exp(\beta)$ was 0.70 to 1.1, corresponding to β from -0.36 to 0.1. Figures 2B and 2C show the estimates of $SCE(t)$ over a much larger range, for β from -3 to 3. The open circles (and ‘+’ and ‘×’ signs) in the plots

represent the sharp bounds of $SCE(t)$ (and 95% Wald and percentile confidence intervals, respectively, for these bounds based on 500 bootstrap replications), corresponding to an analysis with $\beta = \pm\infty$. (Confidence intervals based on the analytic variance were similar.) Regardless of the range, it is clear from the Figure 2 that the vaccine is having no causal effect on the initiation of ART.

It may be of more interest to look at the effect of vaccination on the time-to-event outcomes for the non-white subgroup, where the vaccine appeared to partially protect against HIV infection. These analyses are shown in Figure 3. Because the estimate for VE is much larger in the non-white cohort, the bounds for $F_p^{ai}(t)$ are farther apart, as seen in Figure 3A. This is also reflected in plots B and C, as the estimates for $SCE(t)$ cover a wider range. Of course, the smaller sample size ($N = 914$ non-whites, of which 59 became infected) also inflates the length of the confidence intervals. It should be noted that 95% Wald-based confidence intervals of $SCE(t)$ at $\beta = \infty$ (where $\hat{F}_p^{ai,L}(t) = 1$) using the analytic variance expression were much wider than the bootstrap confidence intervals.

Notice that for these analyses in the non-white cohort, if $\beta > 0.5$ or 0 for $t = 1.17$ or 2 years, respectively, then $H_0 : SCE(t) = 0$ is rejected at the 0.05-level. Let us specifically consider the analysis for $t = 2$ years. This means that given infection in the placebo arm, if the odds of infection if randomized to vaccine are greater for someone who has a longer time to the initiation of ART, then there is evidence that the vaccine is causing non-white participants to have a higher probability of starting ART by 2 years post-infection diagnosis. This would imply that among non-whites the vaccine is having a detrimental effect, causing more rapid post-infection progression. Interestingly, this value of β is just inside Dr. Gurwith's plausible range for non-whites, -0.92 to 0.18 . Therefore, both the null and the alternative are favored in the range. However, it is worth noting that nowhere in this range is the point estimate positive; hence, the analysis provides no support for

the vaccine being effective in non-whites.

6 DISCUSSION

As candidate vaccines continue to be developed and enter the clinical trial phase, there is particular interest in looking at the effect of vaccination on post-infection outcomes. In this paper, we have proposed sensitivity analysis methods for evaluating the causal effect of vaccination on outcomes defined as the time from infection diagnosis to some post-infection event. We discussed non-parametric, parametric, and semiparametric estimation approaches, applying our methods to investigate the causal effect of VaxGen’s AIDS VAX B/B vaccine on the time from infection diagnosis to initiation of ART.

From our VaxGen analysis in the non-white cohort, different conclusions were drawn over the chosen range. To many, the fact that our analyses produce more than one answer may not be attractive. We believe, however, that it is an honest way to present the data and allows scientists to look at results and draw their own conclusions. Based on our experience with Dr. Gurwith, eliciting a range for sensitivity parameters is feasible, although admittedly, different subject matter experts may choose different ranges for β .

It should be noted that the asymptotic normality of these estimators relies on the assumption that $VE > 0$. For VE near the boundary $VE = 0$, Wald-based confidence intervals may have poor coverage (see Jemai and Rotnitzky (submitted, 2005) for further discussion). In the VaxGen trial, using the entire cohort the estimate for VE was 0.048. We performed an additional simulation investigating the performance of Wald-based 95% confidence intervals of $SCE(t = 24 \text{ months})$ under the same settings as described in the Section 5 only generating data under the assumption that $VE = 0.05$ and $\beta = 0.1$. Coverage using the bootstrap and analytic variance estimates was good, 0.941 and 0.938 respectively. Of course, in the VaxGen trial we do not know the true VE so it may be

misleading to read too much into these simulations. This is not an issue for the analysis of the non-white cohort where $\widehat{VE} = 0.469$. However, the asymptotic normality of $\widehat{F}_p^{ai,L}(t)$ for $t = 1.17$ and 2 years could be questioned since $\widehat{VE} > \widehat{F}_p(t)$.

Future research related to this work would therefore be to study methods for creating confidence intervals near the boundaries. Other extensions to this work that may be useful areas of future research include the development of simultaneous confidence bands, construction of a “log-rank” type test, accomodating informative censoring, and semi-parametric methods allowing the inclusion of continuous baseline covariates. Although these methods were designed specifically for HIV vaccine trials, they may apply in other situations where causal comparisons conditioning on post-randomization variables are of interest, so long as the basic assumptions (particularly monotonicity) are reasonable.

ACKNOWLEDGEMENTS

We would like to thank VaxGen for allowing us to use their data, and particularly Marc Gurwith for giving us ranges for the sensitivity parameter.

REFERENCES

- ANDERSEN, P.K., BORGAN, O., GILL, R.D., AND KEIDING, N. (1992). *Statistical Models Based on Counting Processes*. Berlin: Springer.
- CHEN, P.Y. AND TSIATIS, A.A. (2001). Causal inference on the difference of the restricted mean lifetime between two groups *Biometrics* **57**, 1030–1038.
- FLEMING, T.R. AND HARRINGTON, D.P. (1991). *Counting processes and Survival Analysis*. Wiley, New York.

- FLYNN, N.M., FORTHAL, D.N., HARRO, C.D., JUDSON, F.N., MAYER, K.H., PARA, M.F. AND THE RGP120 HIV VACCINE STUDY GROUP (2005). Placebo-controlled phase 3 trial of a recombinant glycoprotein 120 vaccine to prevent HIV-1 infection *Journal of Infectious Diseases* **191**, 647–649.
- FRANGAKIS, C.E. AND RUBIN, D.B. (2002). Principal stratification in causal inference *Biometrics* **58**, 21–29.
- GILBERT, P.B., ACKERS, M.L., BERMAN, P.W., FRANCIS, D.P., POPOVIC, V., HU, D.J., HEYWARD, W.L., SINANGIL, F., SHEPHERD, B.E., GURWITH, M. (2005). HIV-1 virologic and immunologic progression and antiretroviral therapy initiation among HIV-1-infected participants in an efficacy trial of recombinant glycoprotein 120 vaccine. *Journal of Infectious Diseases*, in press.
- GILBERT, P.B., BOSCH, R.J., AND HUDGENS, M.G. (2003). Sensitivity analysis for the assessment of causal vaccine effects on viral load in HIV vaccine trials. *Biometrics* **59**, 531–541.
- GRAHAM, B.S. (2002). Clinical trials of HIV vaccines. *Annual Review of Medicine* **53**, 207–221.
- HALLORAN, M.E. AND STRUCHINER, C.J. (1995). Causal inference for infectious diseases. *Epidemiology* **6**, 142–151.
- HUDGENS, M.G., HOERING, A., AND SELF, S.G. (2003). On the analysis of viral load endpoints in HIV vaccine trials. *Statistics in Medicine* **22**, 2281–2298.
- KALBFLEISH, J.D. AND PRENTICE, R.L. (1980). *The Statistical Analysis of Failure Time Data*. New York: Wiley.

- NABEL, G.J. (2001). Challenges and opportunities for development of an AIDS vaccine. *Nature* **410**, 1002–1007.
- NEYMAN, J. (1923). On the application of probability theory to agricultural experiments: Essay on principles. Translated in *Statistical Science* **5**, 465–480.
- ROBINS, J.M. (1986). A new approach to causal inference in mortality studies with sustained exposure periods - Application to control of the healthy worker survivor effect. *Mathematical Modeling* **7**, 1393–1512.
- ROBINS, J.M. (1995). An analytic method for randomized trials with informative censoring: Part I. *Lifetime Data Analysis* **1**, 241–254.
- ROBINS, J.M. AND GREENLAND, S. (2000). Comment on “Causal inference without counterfactuals,” by A.P. Dawid. *Journal of the American Statistical Association* **95**, 431–435.
- ROSENBAUM, P.R. (1984). The consequences of adjustment for a concomitant variable that has been affected by the treatment. *Journal of the Royal Statistical Society, Series A* **147**, 656–666.
- RUBIN, D.B. (1978). Bayesian inference for causal effects: the role of randomization. *The Annals of Statistics* **6**, 34–58.
- RUBIN, D.B. (2000). Comment on “Causal inference without counterfactuals,” by A.P. Dawid. *Journal of the American Statistical Association* **95**, 435–437.
- SCHARFSTEIN, D.O., ROTNITZKY, A., 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., JEMIAI, Y., AND ROTNITZKY, A. (2005). Sensitivity analyses comparing outcomes only existing in a subset selected post-randomization, conditional on covariates, with application to HIV vaccine trials. *Biometrics*, in press.

STUTE, W. (1995). The central limit theorem under random censorship. *The Annals of Statistics* **23**, 422–439.

VAN DER VAART, A.W. (1998). *Asymptotic Statistics*. Cambridge: Cambridge University Press.

APPENDIX

A.1 Asymptotic Normality Semiparametric Estimator

Suppose that data are collected over a finite interval $[0, \tau]$ where τ is fixed as the sample size $N \rightarrow \infty$. Assume SUTVA, (2.1), A.1, A.2, A.3, $0 < VE < 1$, $p_0 > 0$, and let $t \in [0, \tau]$. Furthermore, assume $0 < w(t; \alpha, \beta) < 1$ for all $t > 0$, α , and β ; is constant for $t > \tau$; and is twice continuously differentiable with respect to α , with a bounded second derivative. Then $\hat{F}_p^{ai}(t)$ is consistent and asymptotically normal.

Proof. The estimates $(\hat{p}_0, \hat{\alpha})$ are the solutions to the estimating equations given by (3.5). It is easily seen that the unique solution to the first equation is $\hat{p}_0 = n_p/N_p$. The second equation can therefore be written as

$$U_N(\alpha) \equiv \int_0^\infty w(t; \alpha, \beta) d\hat{F}_p(t) - (1 - \widehat{VE}) = 0.$$

Using the empirical distribution of O_i , one can show that the process $\hat{F}_p(\cdot)$ in $D[0, \tau]$ and \widehat{VE} are jointly asymptotically normal, where $D[0, \tau]$ is the space of cadlag functions

(i.e., right continuous whose limits from the left exist) on $[0, \tau]$. The map $(\hat{F}_p(\cdot), \widehat{VE}) \mapsto U_N(\alpha)$ is Hadamard differentiable (this follows from Problem 7, Chapter 20, van der Vaart (1998)); therefore, by the functional delta method, $U_N(\alpha)$ is asymptotically normal.

From Taylor's expansion one can write

$$\sqrt{N}(\hat{\alpha} - \alpha) = \frac{-\sqrt{N}U_N(\alpha)}{U'_N(\alpha) + \frac{1}{2}U''_N(\alpha^*)(\hat{\alpha} - \alpha)^2},$$

where α^* is some value between α and $\hat{\alpha}$, and $U''_N(\alpha^*) = \int_0^\infty w''(t; \alpha^*, \beta)d\hat{F}_p(t)$. Notice that $U''_N(\alpha^*)$ is bounded because $|w''(t; \alpha, \beta)|$ is bounded. Therefore, because $U'_N(\alpha) \rightarrow^P U'(\alpha) \equiv \int_0^\infty w'(t; \alpha, \beta)dF_p(t)$ by the continuous mapping theorem, $\hat{\alpha} \rightarrow^P \alpha$ by Lemma 5.10 of van der Vaart (1998), $U''_N(\alpha^*)$ is bounded, and $U_N(\alpha)$ is asymptotically normal; $\hat{\alpha}$ is asymptotically normal. Finally, consider the map:

$$\left(\hat{\alpha}, \hat{F}_p(\cdot)\right) \mapsto \frac{\int_0^t w(s; \hat{\alpha}, \beta)d\hat{F}_p(s)}{\int_0^\infty w(t; \hat{\alpha}, \beta)d\hat{F}_p(t)} \equiv \hat{F}_p^{ai}(t).$$

That the process $\hat{F}_p^{ai}(\cdot)$ in $D[0, \tau]$ is asymptotically normal follows from Hadamard differentiability of this map, the chain rule, and the functional delta method. \square

A.2 Asymptotic Variance of Semiparametric Estimator

From Stute (1995), we learn that $\int \phi(t)d\hat{F}_p(t)$ can be written as a sum of i.i.d. terms plus some remainder term, R_{n_p} , where $|R_{n_p}| = o(n_p^{-1/2})$, and $\phi(t)$ is some well-behaved function of t . Define k as the number of distinct failure times, and let t_1, \dots, t_k represent the distinct ordered failure times. One could think of there being $k + 2$ parameters to estimate, $(p_0, \alpha, F_p(t_1), F_p(t_2), \dots, F_p(t_k)) \equiv \theta$, adding k estimating equations to (3.5):

$$\psi_i(\theta) = \begin{cases} (1 - Z_i)(S_i - p_0) \\ Z_i (S_i - p_0 \int_0^\infty w(s; \alpha, \beta) dF_p(s)) \\ (1 - Z_i)S_i (V_{1i} - F_p(t_1)) \\ \vdots \\ (1 - Z_i)S_i (V_{ki} - F_p(t_k)), \end{cases}$$

with V_{ji} for $j = 1, \dots, k$ and $i = 1, \dots, N$ defined as

$$V_{ji} = \phi_j(Y_i)\gamma_0(Y_i)\delta_i + \gamma_{j1}(Y_i)(1 - \delta_i) - \gamma_{j2}(Y_i), \text{ where}$$

$$\phi_j(Y_i) = I_{\{Y_i \leq t_j\}},$$

$$\gamma_0(Y_i) = \exp\left(\int_{-\infty}^{Y_i^-} \frac{H^0(dz)}{1 - H(z)}\right),$$

$$\gamma_{j1}(Y_i) = \frac{1}{1 - H(Y_i)} \int I_{\{Y_i < \omega\}} \phi_j(\omega)\gamma_0(\omega)H^1(d\omega),$$

$$\gamma_{j2}(Y_i) = \int \int \frac{I_{\{\nu < Y_i, \nu < \omega\}} \phi_j(\omega)\gamma_0(\omega)}{(1 - H(\nu))^2} H^0(d\nu)H^1(d\omega),$$

$$H^0(y) = \mathbb{P}_N(Y \leq y, \delta = 0) = \frac{\sum (1 - z_i) s_i (1 - \delta_i) I_{\{Y_i \leq y\}}}{\sum (1 - z_i) s_i},$$

$$H^1(y) = \mathbb{P}_N(Y \leq y, \delta = 1) = \frac{\sum (1 - z_i) s_i \delta_i I_{\{Y_i \leq y\}}}{\sum (1 - z_i) s_i},$$

$$H(y) = \mathbb{P}_N(Y \leq y) = \frac{\sum (1 - z_i) s_i I_{\{Y_i \leq y\}}}{\sum (1 - z_i) s_i}.$$

To be clear and to simplify further notation:

$$\int_0^\infty w(s; \alpha, \beta) dF_p(s) = \sum_{j=1}^{k+1} w_j(\alpha) (F_p(t_j) - F_p(t_{j-1})),$$

where $w_j(\alpha) = w(t_j; \alpha, \beta)$, $w_{k+1}(\alpha) = w(\tau; \alpha, \beta)$, $F_p(t_0) = 0$, and $F_p(t_{k+1}) = 1$.

From Appendix A.2 we know that $\sqrt{N}(\hat{\theta} - \theta) \rightarrow^d \mathcal{N}(0, \Psi)$, where

$$\Psi = E \left[\frac{\partial}{\partial \theta} \psi(\theta) \right]^{-1} E [\psi(\theta) \psi(\theta)^T] E \left[\left(\frac{\partial}{\partial \theta} \psi(\theta) \right)^T \right]^{-1},$$

Define

$$g(\theta) = \frac{\sum_{j=1}^l w_j(\alpha)(F_p(t_j) - F_p(t_{j-1}))}{\sum_{j=1}^{k+1} w_j(\alpha)(F_p(t_j) - F_p(t_{j-1}))},$$

where $t_l = \sup(t_j)$ such that $t_j < t$. Therefore $g(\hat{\theta}) = \hat{F}_p^{ai}(t)$. By the delta method,

$$\sqrt{N} \left(g(\hat{\theta}) - g(\theta) \right) \rightarrow^d \mathcal{N} \left(0, g'(\theta) \Psi g'(\theta)^T \right).$$

One may estimate $g'(\theta)$ with $g'(\hat{\theta})$ and Ψ with

$$\hat{\Psi} = \left[\frac{1}{N} \sum_{i=1}^N \frac{\partial}{\partial \theta} \psi_i(\hat{\theta}) \right]^{-1} \frac{1}{N} \sum_{i=1}^N \left[\psi_i(\hat{\theta}) \psi_i(\hat{\theta})^T \right] \left[\frac{1}{N} \sum_{i=1}^N \left(\frac{\partial}{\partial \theta} \psi_i(\hat{\theta}) \right)^T \right]^{-1}.$$

Table 1: Bias of estimates and coverage probability of Wald-based 95% confidence intervals for $SCE(t)$ with $t = 24$ months.

VE	β	Bias		Coverage Probability	
		Mean	Median	Bootstrap	Analytic
~ 0.3	0.1	-0.002	0.000	0.946	0.948
	0.2	-0.007	-0.007	0.943	0.949
	1	-0.006	-0.003	0.943	0.946
	∞	-0.007	0.005	0.948	0.953
~ 0.6	0.1	0.003	0.006	0.939	0.940
	0.2	0.013	0.015	0.945	0.945
	1	0.042	0.020	0.933	0.912
	∞	0.035	0.000	0.935	–

Analytic coverage probabilities when $VE \approx 0.6$, $\beta = \infty$ are not given because in many simulations, there were no failures in the vaccine arm.

Table 2: Bias of estimates and coverage probability of Wald-based 95% confidence intervals for $F_p^{ai}(t)$ with $t = 24$ months.

VE	β	Bias		Coverage Probability			
		Mean	Median	Standard Bootstrap	Standard Analytic	Log-log transformation Bootstrap	Log-log transformation Analytic
~ 0.3	0.1	-0.001	-0.002	0.939	0.941	0.946	0.942
	0.2	-0.006	-0.002	0.932	0.939	0.938	0.949
	1	0.001	-0.005	0.940	0.953	0.959	0.973
	∞	-0.007	0.002	0.934	0.947	0.957	0.967
~ 0.6	0.1	0.001	0.003	0.939	0.942	0.947	0.947
	0.2	0.015	0.011	0.929	0.927	0.938	0.937
	1	0.045	0.010	0.903	0.819	0.902	0.859
	∞	0.036	-0.016	0.889	0.948	–	–

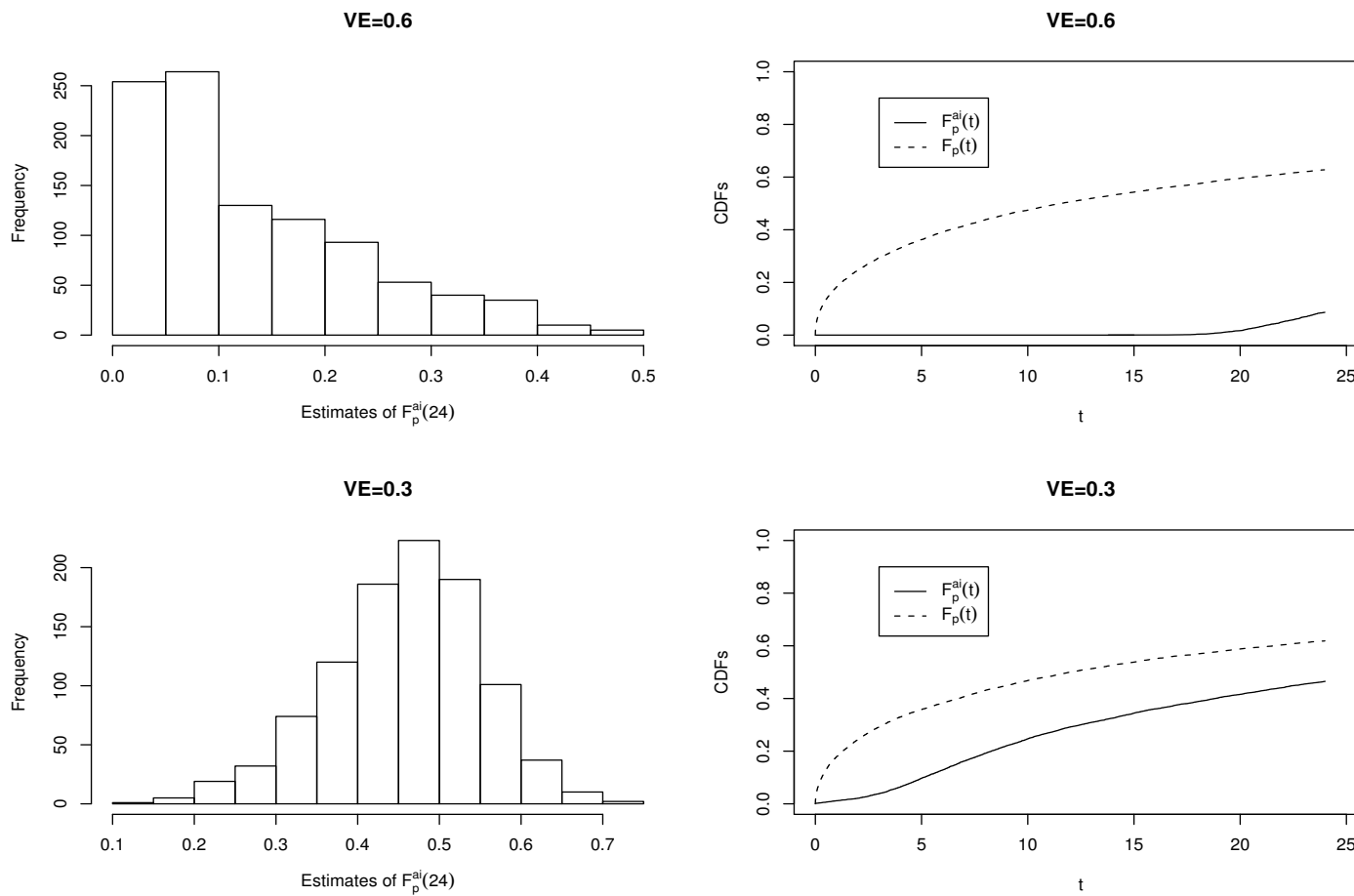


Figure 1: Histograms of $\hat{F}_p^{ai}(t)$ for $t = 24$ and plots of $F_p(t)$ and $F_p^{ai}(t)$ at different levels of VE for $\beta = 1$.

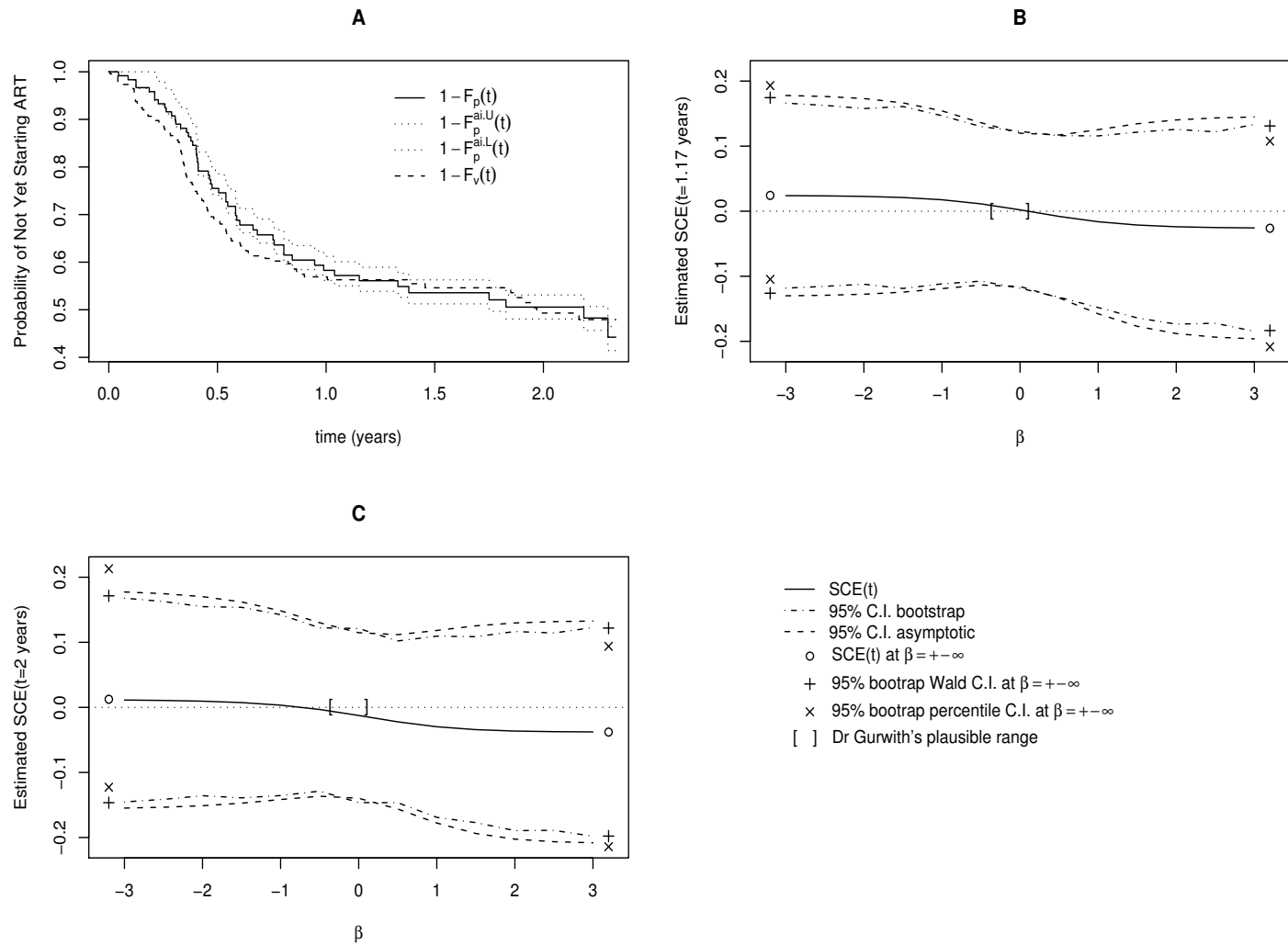


Figure 2: Sensitivity analyses of the causal effect of vaccination on the probability of not yet initiating ART.

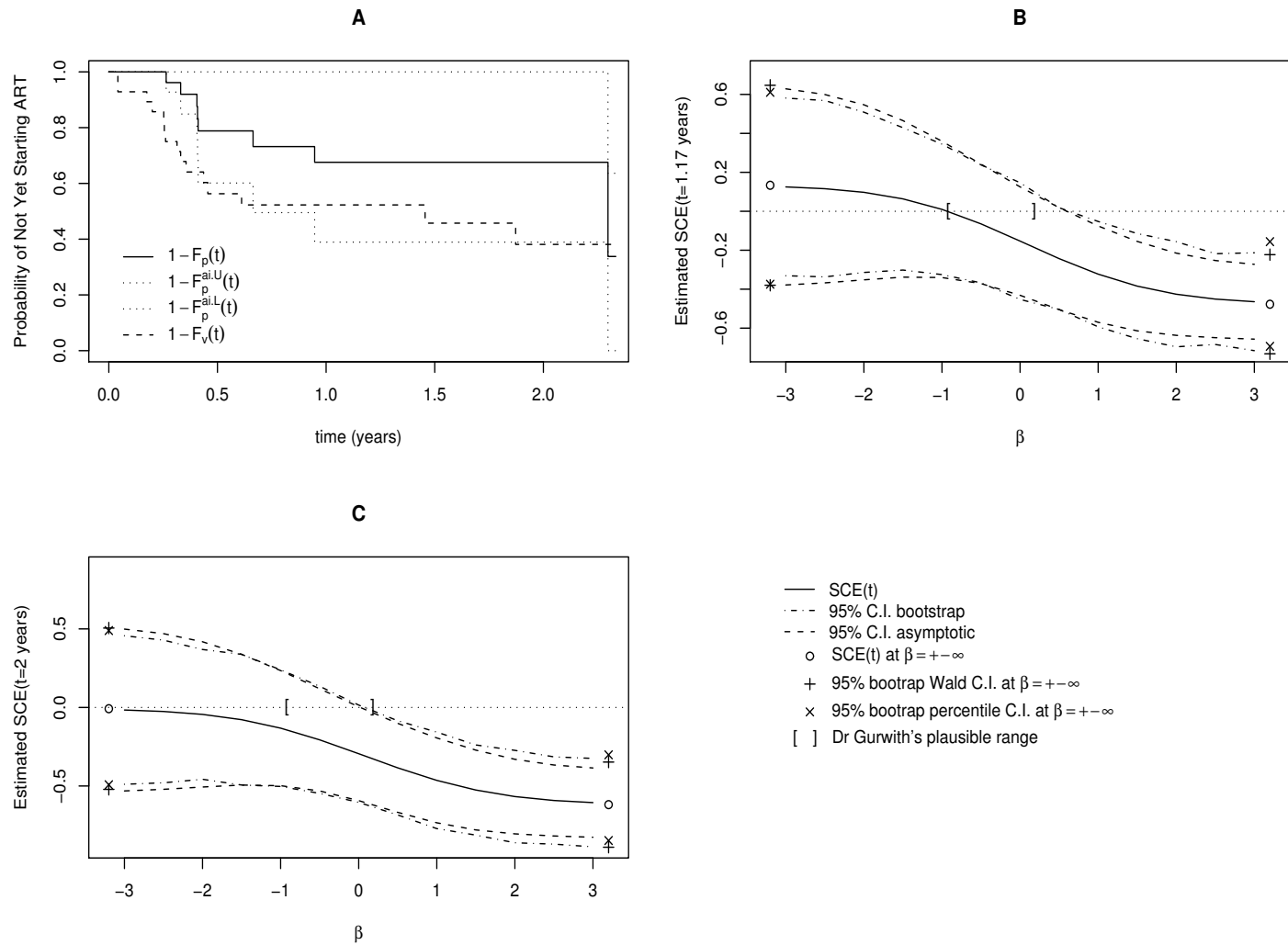


Figure 3: Sensitivity analyses of the causal effect of vaccination on the probability of not yet initiating ART in the non-white cohort.