

# Sensitivity Analyses Comparing Outcomes Only Existing in a Subset Selected Post-Randomization, Conditional on Covariates, with Application to HIV Vaccine Trials

Bryan E. Shepherd,<sup>1,\*</sup> Peter B. Gilbert,<sup>1</sup> Yannis Jemai,<sup>2</sup> and Andrea Rotnitzky<sup>2</sup>

<sup>1</sup>Department of Biostatistics, University of Washington

<sup>2</sup>Department of Biostatistics, Harvard School of Public Health

\**email*: bshepher@u.washington.edu

**SUMMARY.** In many experiments researchers would like to compare between treatments an outcome that only exists in a subset of participants selected after randomization. For example, in preventive HIV vaccine efficacy trials it is of interest to determine whether randomization to vaccine causes lower HIV viral load, a quantity that only exists in participants who acquire HIV. To make a causal comparison and account for potential selection bias we propose a sensitivity analysis following the principal stratification framework set forth by Frangakis and Rubin (2002). Our goal is to assess the average causal effect of treatment assignment on viral load at a given baseline covariate level in the always infected principal stratum (those who would have been infected whether they had been assigned to vaccine or placebo). We assume stable unit treatment values (SUTVA), randomization, and that subjects randomized to the vaccine arm who became infected would also have become infected if randomized to the placebo arm (monotonicity). It is not known which of those subjects infected in the placebo arm are in the always infected principal stratum, but this can be modeled conditional on covariates, the observed viral load, and a specified sensitivity parameter. Under parametric regression models for viral load, we obtain maximum likelihood estimates of the average causal effect conditional on covariates and the sensitivity parameter. We apply our methods to the world's first Phase III HIV vaccine trial.

**KEY WORDS:** Causal inference; principal stratification; maximum likelihood; viral load.

## 1. Introduction

Vaccines are being developed in an attempt to curb the spread of HIV: the first Phase III preventive HIV vaccine trial was recently completed (VaxGen's trial of AIDSVAX B/B reported in Flynn et al., 2005), other efficacy trials are in progress, and many more are expected. Of primary interest in these trials is whether the vaccine protects against HIV infection; a perfect vaccine would eliminate infection. However, a less than perfect vaccine could also reduce morbidity and mortality by preventing infection for some and ameliorating disease progression or decreasing infectiousness among those who acquire HIV (Nabel, 2001; Graham, 2002).

To that end, investigators are usually interested in the effect of vaccine on HIV viral load, a commonly used surrogate variable for measuring the extent of an infected individual's HIV disease and infectiousness (O'Brien et al., 1996; Quinn et al., 2000). Two types of questions emerge from such investigations. Clinicians or epidemiologists are ultimately interested in whether they should recommend vaccine to the general population (or a targeted subgroup therein). Scientists, however, may want to know whether there exists a mechanism through which the vaccine alters viral load in infected individuals, perhaps leading to further scientific innovation and the elaboration of new vaccines. The methods in this article are geared towards answering questions of the second type. They only have secondary utility in answering questions of the first type. An intention to treat analysis (ITT) with a loss function assigning appropriate weights to infection and viral load after infection would be the correct primary analysis from which to draw inference for policy recommendations (Robins and Greenland (2000); Gilbert, Bosch, and Hudgens (2003) (GBH)). In this article, we will incorporate baseline covariate information into the analysis in order to refine the scientist's understanding of the mechanism by which the vaccine acts, specifically of the modifiers of the vaccine effect after infection.

Naïvely comparing the distribution of viral loads between HIV infected individuals in the vaccine and placebo arms of a randomized experiment within levels of the baseline covariates, does not lead to unbiased estimates of the causal vaccine effect on viral load after infection because it improperly conditions on a post-randomization variable, HIV infection (Rosenbaum, 1984; Halloran and Struchiner, 1995). An alternative analytical strategy would assign a viral load value of 0 (or best rank) to all uninfected participants and perform an ITT analysis. This analysis addresses a causal question, but as verified in simulations, will lack power because most of the vaccine effect is washed out by the zeroes, which can occur in a large fraction of participants in an HIV vaccine trial (93% for the VaxGen trial).

As pointed out by many authors (Kalbfleish and Prentice 1980; Robins 1995; Rubin 2000; Robins and Greenland 2000), a meaningful causal effect on viral load is defined in the subset of the population which would become infected under either placebo or vaccine. Specifically, each subject has a potential infection status if assigned vaccine and a potential infection status if assigned placebo, only one of which is observed during the trial. In addition, every subject that would be infected under a treatment (vaccine or placebo) also has a potential viral load under that treatment. Every subject can be classified into one of four possible combinations of the two potential infection status outcomes: never infected (not infected if assigned vaccine or placebo), harmed (infected if assigned vaccine but not infected if assigned placebo), protected (not infected if assigned vaccine but infected if assigned placebo), and always infected (infected regardless of assignment). This classification has been referred to as principal stratification by Frangakis and Rubin (2002). Only in the always infected ( $ai$ ) principal stratum do subjects have a potential viral load under both treatments. Therefore, only in the  $ai$  principal stratum are causal comparisons meaningful. Thus the type of questions addressed in this paper are whether

in individuals with given characteristics (i.e. young caucasian males with no history of venereal disease who would become infected regardless of vaccine/placebo assignment), the vaccine alters viral load. The vaccine trial cannot provide a definite answer to such questions because an individual's principal stratum membership can never be known and thus the distribution of viral load in the  $ai$  stratum under each treatment is not identified. One goal of this paper is to derive a set of reasonable and easily understandable assumptions under which such distributions are identified by the clinical trial data. A second goal is to indicate how to conduct inference about the treatment effect in the  $ai$  stratum under such assumptions.

Hudgens, Hoering, and Self (2003) (HHS) and GBH addressed the same question in the absence of baseline covariates. These authors made the monotonicity assumption which postulates that an individual who would get infected under vaccine would also get infected under placebo. Monotonicity is possibly not an unrealistic assumption when comparing vaccine to placebo. It implies that the vaccine effect on infection risk is either beneficial or harmless, and that all individuals infected in the vaccine group belong to the  $ai$  principal stratum. Thus, monotonicity is sufficient to identify the viral load distribution under vaccine in the  $ai$  stratum. However, monotonicity is not sufficient to identify the viral load distribution under placebo in the  $ai$  stratum because infected individuals randomized to placebo can belong to either the  $ai$  or the protected principal strata. To identify this distribution, GBH assumed that the probability that a subject who becomes infected under placebo is in the  $ai$  stratum depends on his/her viral load  $y$  under placebo through the expit function  $w(y) = e^{\alpha+\beta y} / (1 + e^{\alpha+\beta y})$ . GBH showed that the parameter  $\beta$  is not identified by the clinical trial data but once it is specified, the distribution of the viral load under placebo in the  $ai$  stratum and  $\alpha$  are both identified. Thus, similar to Scharfstein, Rotnitzky, and Robins (1999), GBH advocated carrying out a sensitivity

analysis in which  $\beta$  is varied along a plausible range and inference about vaccine effects in the  $ai$  stratum is repeated for each value of  $\beta$  in the range. If results hold in one direction for a plausible range of the sensitivity parameter  $\beta$ , then a causal conclusion in that direction may be drawn. Otherwise, the analysis remains inconclusive.

In this article, we extend the approach of GBH by describing sensitivity analysis methods for estimating treatment effects conditional on baseline covariates. We first define assumptions which identify the causal estimand of interest and then describe parametric models under which their estimation is feasible. Next, we derive the likelihood under the assumed parametric models and indicate how maximum likelihood estimates can be obtained. Finally, we exhibit results from simulation studies of our estimators' finite sample properties, and apply our methods to data from the VaxGen HIV vaccine trial.

## 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 recorded data on subject  $i$  are a vector of baseline covariates  $X_i$ , an indicator  $S_i$  of infection during the study follow-up period ( $S_i = 1$  if infected and  $S_i = 0$  if not), and, if infected, the viral load  $Y_i$  (on a  $\log_{10}$  scale) shortly after diagnosis of infection.

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 indicator if, possibly contrary to fact, subject  $i$  is assigned placebo. Define  $S_i(1)$  to be the infection indicator if subject  $i$  is assigned vaccine. Similarly, define  $Y_i(0)$  to be the viral load if participant  $i$  is assigned placebo and  $Y_i(1)$  the viral load if assigned vaccine,

where for a subject who does not become infected, i.e.,  $S_i(j) = 0$ , we define  $Y_i(j) = *$ ,  $j = 0, 1$ . The notation implicitly assumes that the potential outcomes of each trial participant are not influenced by the treatments of other participants, known as the Stable Unit Treatment Value Assumption (SUTVA) (Rubin, 1978, 1986). It implies consistency,

$$Y_i(j) = Y_i \text{ if } Z_i = j, \quad j = 0, 1. \quad (1)$$

Assuming the study participants make up a random sample from a large population of interest, then the outcomes  $W_i = (S_i(0), S_i(1), Y_i(0), Y_i(1), Z_i, X_i)$ ,  $i = 1, \dots, N$ , are i.i.d. copies of a random vector  $W = (S(0), S(1), Y(0), Y(1), Z, X)$ , and similarly the observed data  $O_i = (Z_i, X_i, S_i, Y_i)$ ,  $i = 1, \dots, N$ , are i.i.d. copies of  $O \equiv (Z, X, S, Y)$ , where we define  $Y = *$  if  $S = 0$ .

Randomization, possibly depending on the baseline covariates  $X$ , ensures that

$$(S(0), S(1), Y(0), Y(1)) \perp\!\!\!\perp Z | X \quad (2)$$

because  $(S(0), S(1), Y(0), Y(1))$  can, like genetic make-up, be considered an unobserved baseline characteristic of each subject. Here, for random variables  $A, B$  and  $C$ ,  $A \perp\!\!\!\perp B | C$  indicates conditional independence of  $A$  and  $B$  given  $C$ .

The four principal strata described in the introduction 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 are those with  $S(0) = S(1) = 1$ .

For a subject  $i$  who is in the  $ai$  principal stratum, a causal effect measure on his/her viral load is some measure of discrepancy between  $Y_i(0)$  and  $Y_i(1)$ , for example the difference  $Y_i(1) - Y_i(0)$ . The average causal effect at covariate level  $X = x$  in the  $ai$  stratum is defined as

$$ACE(x) = E(Y(1) - Y(0) | S(0) = S(1) = 1, X = x).$$

Our goal is to propose methods for estimating the function  $ACE(\cdot)$  based on the observed data  $O_i, i = 1, \dots, N$ . Since  $ACE(x)$  is a comparison of conditional means in the always infected principal stratum and because randomization alone does not suffice to determine membership to the  $ai$  stratum, we must make additional assumptions on the distribution of  $W$  in order to identify  $ACE(x)$ . Arguing as in GBH, it can be shown that the following assumptions, together with (1) and (2), do identify  $ACE(x)$ .

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

A.2: A model for the mixing probabilities of the always infected and protected in the infected placebo group. Specifically,

$$P(S(1) = 1 | S(0) = 1, Y(0), X) = w(X, Y(0); \beta); \quad (3)$$

where  $w(x, y; \beta) = \Phi\{m(x) + g(x, y; \beta)\}$ ,  $\beta$  is fixed and known,  $\Phi(\cdot)$  is a known cdf,  $m(\cdot)$  is an unspecified function of  $X$ , and for each  $\beta$ ,  $g(\cdot, \cdot; \beta)$  is a known function of  $X$  and  $Y$ .

The parameter  $\beta$  is not identified by the observed data. We propose regarding  $\beta$  (and therefore, the function  $g(\cdot, \cdot; \beta)$ ) as fixed and known and, as a form of sensitivity analysis, estimating  $ACE(x)$  under different values of  $\beta$ . The range for  $\beta$  should be chosen independent from the data. Choosing  $\beta$  such that  $g(X, Y(0); \beta) = 0$  is the same as assuming that  $Y(0) \perp\!\!\!\perp S(1) | S(0) = 1, X$  (here referred to as assumption A.3), or equivalently that the distribution of viral loads under placebo is the same in the  $ai$  and protected strata.

If  $X$  is discrete and can take a small number of values, we can estimate  $ACE(x)$  by applying the methods of GBH within each level  $x$ . However, if the distribution of  $X$  is continuous or discrete with a large support this approach is unfeasible because the data are too sparse to conduct cell-specific estimation. We address this problem by imposing the following additional distributional assumptions on the law of  $W$ :

M.1: The probability of infection given covariates when assigned placebo is known up to a finite dimensional parameter  $\mu$ ; that is,  $P(S(0) = 1|X) = \theta_p(X; \mu)$ , where  $\mu$  is unknown and for each  $\mu$ ,  $\theta_p(\cdot; \mu)$  is a known function.

M.2: The function  $m(x)$  in A.2 follows a parametric model  $m(X) = m(X; \alpha)$ , where  $\alpha$  is an unknown parameter vector and for each  $\alpha$ ,  $m(\cdot; \alpha)$  is a known function.

M.3: The conditional distribution of viral load under vaccine given covariates  $X$  in the *ai* stratum is known up to a finite dimensional parameter  $\eta_1$ ; that is,  $f_{Y(1)|S(1)=1, X}(y|S(1) = 1, x) = f_v(y|x; \eta_1)$ , where  $\eta_1$  is unknown and for each  $\eta_1$ ,  $f_v(\cdot|x; \eta_1)$  is a known density.

We also make one of the following two assumptions:

M.4a: The conditional distribution of viral load under placebo given covariates  $X$  in the population comprised of both protected and always infected individuals, is known up to a finite dimensional parameter  $\eta_0^a$ ; that is,  $f_{Y(0)|S(0)=1, X}(y|S(0) = 1, x) = f_p(y|x; \eta_0^a)$ , where  $\eta_0^a$  is unknown and for each  $\eta_0^a$ ,  $f_p(\cdot|x; \eta_0^a)$  is a known density.

M.4b: The conditional distribution of viral load under placebo given covariates  $X$  in the *ai* principal stratum is known up to a finite dimensional parameter  $\eta_0^b$ ; that is,  $f_{Y(0)|S(0)=S(1)=1, X}(y|S(0) = S(1) = 1, x) = f_p^{ai}(y|x; \eta_0^b)$ , where  $\eta_0^b$  is an unknown parameter and for each  $\eta_0^b$ ,  $f_p^{ai}(\cdot|x; \eta_0^b)$  is a known density.

For ease of reference, we call the model defined by assumptions (1), (2), A.1, A.2, M.1-M.3 and M.4a, model  $\mathcal{M}_a$ . We call  $\mathcal{M}_b$  the model defined like  $\mathcal{M}_a$  except that M.4a is replaced by M.4b and in A.2 we demand that  $w(x, y; \beta, \alpha) > 0$  for all  $(x, y, \beta, \alpha)$ . There are advantages and disadvantages to both models, which we discuss in Section 4.

If in assumption A.2 the parameter  $\beta$  of model (3) is regarded as unknown, it is not identified under (1), (2), and assumptions A.1 and A.2. However,  $\beta$  is identified if the distributional forms imposed by M.1-M.3 and M.4b are also assumed. Consequently,



under  $\mathcal{M}_b$  rather than regarding  $\beta$  as fixed and known, one could estimate it. However, since  $\beta$  is only identified because of models imposed to reduce the data dimensionality, rather than estimating it we recommend continuing to regard  $\beta$  as fixed and known and conducting sensitivity analyses over plausible values of  $\beta$ .

### 3. Maximum likelihood estimation of $ACE(x)$ .

Under model  $\mathcal{M}_a$ ,  $ACE(x)$  is a function of the unknown parameters  $(\alpha, \eta_1, \eta_0^a)$ . Specifically,  $ACE(x) = ACE_a(x; \alpha, \eta_1, \eta_0^a)$  where

$$ACE_a(x; \alpha, \eta_1, \eta_0^a) \equiv \int y f_v(y|x; \eta_1) dy - \frac{\int y w(x, y; \beta, \alpha) f_p(y|x; \eta_0^a) dy}{\int w(x, y; \beta, \alpha) f_p(y|x; \eta_0^a) dy}.$$

Similarly, under model  $\mathcal{M}_b$ ,  $ACE(x)$  is a function of  $(\eta_1, \eta_0^b)$  since it is equal to  $ACE_b(x; \eta_1, \eta_0^b) \equiv \int y f_v(y|x; \eta_1) dy - \int y f_p^{ai}(y|x; \eta_0^b) dy$ . The maximum likelihood estimators of  $ACE(x)$  under models  $\mathcal{M}_a$  and  $\mathcal{M}_b$  are therefore equal to the functions  $ACE_a(x; \cdot, \cdot, \cdot)$  and  $ACE_b(x; \cdot, \cdot)$  evaluated at the ML estimators of  $(\alpha, \eta_1, \eta_0^a)$  and  $(\eta_1, \eta_0^b)$ , respectively.

To derive the ML estimator we express the joint density of the observables  $O$ ,

$$f_O(O) = f_X(X) P_{Z|X}(Z|X) P_{S|Z,X}(S|Z, X) f_{Y|S,Z,X}(Y|S, Z, X),$$

in terms of the model parameters. Specifically, in the Appendix we show that

$$f_{Y|S,Z,X}(y|S=1, Z=0, X=x) = \begin{cases} f_p(y|x; \eta_0^a) & \text{under } \mathcal{M}_a \\ f_p^*(y|x; \alpha, \eta_0^b) \equiv \frac{w^{-1}(x, y; \beta, \alpha) f_p^{ai}(y|x; \eta_0^b)}{\int w^{-1}(x, y; \beta, \alpha) f_p^{ai}(y|x; \eta_0^b) dy} & \text{under } \mathcal{M}_b, \end{cases}$$

$$f_{Y|S,Z,X}(y|S=1, Z=1, X=x) = f_v(y|x; \eta_1) \quad \text{under } \mathcal{M}_a \text{ or } \mathcal{M}_b,$$

$$P_{S|Z,X}(S=1|Z=0, X=x) = \theta_p(x; \mu) \quad \text{under } \mathcal{M}_a \text{ or } \mathcal{M}_b,$$

$$P_{S|Z,X}(S=1|Z=1, X=x) = \begin{cases} \theta_p(x; \mu) \int w(x, y; \beta, \alpha) f_p(y|x; \eta_0^a) dy & \text{under } \mathcal{M}_a \\ \theta_p(x; \mu) \int w(x, y; \beta, \alpha) f_p^*(y|x; \alpha, \eta_0^b) dy & \text{under } \mathcal{M}_b. \end{cases}$$

It follows that the likelihoods  $\mathcal{L}_a(\rho^a)$  and  $\mathcal{L}_b(\rho^b)$  for  $\rho^a \equiv (\mu, \alpha, \eta_1, \eta_0^a)$  and  $\rho^b \equiv (\mu, \alpha, \eta_1, \eta_0^b)$  under models  $\mathcal{M}_a$  and  $\mathcal{M}_b$  respectively, with  $\beta$  known, are

$$\begin{aligned} \mathcal{L}_a(\rho^a) &\propto \prod_{i=1}^N \left[ \left\{ f_v(y_i|x_i; \eta_1) \theta_p(x_i; \mu) \int w(x_i, y; \beta, \alpha) f_p(y|x_i; \eta_0^a) dy \right\}^{S_i} \right. \\ &\quad \times \left. \left\{ 1 - \theta_p(x_i; \mu) \int w(x_i, y; \beta, \alpha) f_p(y|x_i; \eta_0^a) dy \right\}^{1-S_i} \right]^{Z_i} \\ &\quad \times \left[ \left\{ \theta_p(x_i; \mu) f_p(y_i|x_i; \eta_0^a) \right\}^{S_i} \left\{ 1 - \theta_p(x_i; \mu) \right\}^{1-S_i} \right]^{1-Z_i}, \end{aligned} \quad (4)$$

and  $\mathcal{L}_b(\rho^b)$  is defined like  $\mathcal{L}_a(\rho^a)$  but with  $f_p^*(y|x; \alpha, \eta_0^b)$  replacing  $f_p(y|x; \eta_0^a)$ .

In some datasets, including that from the VaxGen trial, the outcomes  $Y_i$  may be censored either above or below certain detection limits. The likelihood in such cases can be easily extended to accommodate censored observations. In addition, when  $w(\cdot)$  depends on  $y$ , obtaining the MLEs of  $\rho^a$  and  $\rho^b$  may require maximizing over an integral that is not in closed form. In Section 5 we discuss the numerical methods that we used to maximize the likelihood in our simulations and example.

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  $ACE(x)$  for each fixed  $x$ . In our simulation studies, variance estimates based on the observed and expected information yielded nearly identical inferences. Alternatively, variances may be estimated by bootstrapping.

It is interesting to note that in the absence of baseline covariates with M.3 and M.4 left unspecified, models  $\mathcal{M}_a$  and  $\mathcal{M}_b$  are the same model and the ML estimator of  $ACE \equiv E(Y(1) - Y(0)|S(0) = S(1) = 1)$  coincides with the estimator of  $ACE$  derived

by GBH (Shepherd, unpublished Ph.D. dissertation).

#### 4. Specific Parameterizations

For our simulations and example, under both models  $\mathcal{M}_a$  and  $\mathcal{M}_b$  we have considered

$$\theta_p(x; \mu) = \frac{\exp(x^T \mu)}{1 + \exp(x^T \mu)} \quad (5)$$

$$w(x, y; \beta, \alpha) = \frac{\exp(x^T \alpha + \beta y)}{1 + \exp(x^T \alpha + \beta y)} \quad (6)$$

where  $x = (1, x_1, \dots, x_q)^T$  and  $\mu$  and  $\alpha$  are parameter vectors of length  $q + 1$ .

##### 4.1. Parameterization for model $\mathcal{M}_a$

In order to ensure that the global null hypothesis,  $H_0^{global} : ACE(x) = 0$  for all  $x$ , is not a-priori excluded under model  $\mathcal{M}_a$ ,  $f_v(y|x; \eta_1)$  in M.3 must have the functional form

$$f_v(y|x; \eta_1) = \frac{w(x, y; \beta_v, \alpha_v) f_p(y|x; \eta_v)}{\int w(x, y; \beta_v, \alpha_v) f_p(y|x; \eta_v) dy}, \quad (7)$$

where  $\eta_1 = (\alpha_v, \beta_v, \eta_v)$  is an unknown parameter.

Model  $\mathcal{M}_a$  has two drawbacks. First, natural functional forms for  $ACE(x)$  when  $H_0$  fails cannot be expressed as simple restrictions on the parameters of the model. Second, because of the many parameters in (7) there are identification problems. Specifically, at the value of  $\beta_v$  that makes  $w(x, y; \beta_v, \alpha_v)$  a function of  $x$  only, the parameter  $\alpha_v$  is not identified. For example, under (6) the parameter  $\alpha_v$  is not identified at  $\beta_v = 0$ . Also, depending on the functional forms of  $w(x, y; \cdot, \cdot)$  and  $f_p(y|x; \cdot)$ , there may be additional identification problems. For example if, as we assume in our example and simulations,

$$f_p(y|x; \eta_0^a) = \phi(y; x^T \lambda, \sigma^2) \quad (8)$$

where  $\eta_0^a = (\lambda, \sigma)$ ,  $\lambda = (\lambda_0, \lambda_1, \dots, \lambda_q)^T$  and  $\phi(y; x^T \lambda, \sigma^2)$  is a normal density with mean  $x^T \lambda$  and variance  $\sigma^2$ , then with  $\eta_v = (\lambda_v, \sigma_v)$ , all values of  $\eta_1 = (\alpha_v, \beta_v, \lambda_v, \sigma_v)$  that satisfy

$\alpha_v - (-\beta_v \lambda_{v0} - \beta_v^2 \sigma_v^2 / 2, -\beta_v \lambda_{v1}, \dots, -\beta_v \lambda_{vq}) = 0$  give the same function  $f_v(y|x; \eta_1)$ . That  $\eta_1$  is not identifiable may not appear to be a problem because it does not rule out the identifiability of  $f_v(y|x; \eta_1)$  and hence of  $E(Y(1)|S(1) = 1, X = x)$  which is the relevant term needed to compute  $ACE(x)$ . However, the lack of identification of  $\eta_1$  does complicate inference. Specifically, standard theory does not apply for the asymptotic distribution of ML estimators of  $\eta_1$ . Consequently, using the delta method with the usual calculations for the asymptotic variance of the ML estimator of  $\eta_1$  to estimate the variance of the ML estimator of  $ACE(x)$  yields Wald confidence intervals whose coverage probability, even with large samples, is not close to the nominal level. Even though standard asymptotic calculations do not result in consistent variance estimators, we suspect that bootstrap estimators of the variance of the estimate of  $E(Y(1)|S(1) = 1, X = x)$  can be used to compute valid confidence intervals. Our suspicion follows from the conjecture that the ML estimators of  $E(Y(1)|S(1) = 1, X = x)$  may be regular and asymptotically normal even when  $\eta_1$  is not identified. Our simulations support this conjecture, although this warrants further study.

#### 4.2. Parameterization for model $\mathcal{M}_b$

Under  $\mathcal{M}_b$ ,  $f_{Y(0)|S(0)=S(1)=1, X}(y|S(0) = S(1) = 1, x)$  is modeled directly. A major advantage of this model is that, as opposed to model  $\mathcal{M}_a$ , the functional form for  $f_v(y|x; \eta_1)$  can be easily chosen so as to ensure both that the global null hypothesis can hold and that  $\eta_1$  is identified and  $\sqrt{n}$ -estimable. In our example and simulations we use

$$f_p^{ai}(y|x; \eta_0^b) = \phi(y; x^T \gamma_p, \sigma_p^2) \text{ and } f_v(y|x; \eta_1) = \phi(y; x^T \gamma_v, \sigma_v^2), \quad (9)$$

where  $\eta_0^b = (\gamma_p, \sigma_p)$ ,  $\eta_1 = (\gamma_v, \sigma_v)$ ,  $\gamma_p = (\gamma_{p0}, \gamma_{p1}, \dots, \gamma_{pq})^T$ ,  $\phi(y; x^T \gamma_v, \sigma_v^2)$  is a normal density with mean  $x^T \gamma_v$  and variance  $\sigma_v^2$ , and  $\gamma_p, \sigma_p, \gamma_v$ , and  $\sigma_v$  are unknown. Note that under (9),  $ACE(x)$  takes the simple and easily interpretable form  $x^T (\gamma_v - \gamma_p)$ . Thus, for example, a test of  $H_0^{inter} : \gamma_{vj} = \gamma_{pj}$  is a test of no interaction between treatment and

the  $j$ th covariate, and a global test of the average causal effect of vaccine on viral load is a test of  $H_0^{global} : \gamma_v = \gamma_p$ . Interestingly, under (6) and (9) a closed form expression for the likelihood  $\mathcal{L}_b(\rho^b)$  exists because  $\int w(x, y; \beta, \alpha) f_p^*(y|x; \alpha, \eta_0^b) dy = [1 + \exp\{-\beta(x^T \gamma_p - \frac{1}{2}\beta\sigma_p^2) - x^T \alpha\}]^{-1}$ .

One potential disadvantage of model  $\mathcal{M}_b$  is that it requires that  $w(x, y; \beta, \alpha)$  is nonzero for all  $x$  and  $y$ . This rules out sharp bound analyses (HHS) and is tantamount to assuming the vaccine is not 100% effective in any subpopulation. However, this is a plausible assumption for most candidate HIV vaccines (see Graham, 2002), and even if there were subpopulations where the vaccine was known to be 100% effective, participants in these subpopulations would not belong to the  $ai$  principal stratum and could therefore simply be removed before performing the analysis.

Note that by assuming a functional form for  $f_{Y(0)|S(0)=S(1)=1, X}(y|S(0) = S(1) = 1, x)$  and another for  $P(S(1) = 1|S(0) = 1, Y(0), X = x)$  we are indirectly imposing functional form restrictions on the distribution  $f_{Y(0)|S(0)=1, X}(y|S(0) = 1, x)$ , which is identified by the observed data without assumptions A.1 and A.2. Indeed, this functional form is strongly driven by the chosen value of  $\beta$ . This is not surprising since, as argued earlier,  $\beta$  is identified under model  $\mathcal{M}_b$ . A consequence of this remark is that some choices of  $\beta$  can result in poor model fits. For example, extreme values of  $\beta$  may correspond to a bimodal distribution for  $f_{Y(0)|S(0)=1, X}(y|S(0) = 1, x)$  and this may be contradicted by the evidence in the data. If such extreme values of  $\beta$  were indeed regarded as plausible prior to assuming M.1-M.4b, then we recommend that the analyst consider more flexible distributional shape assumptions since poor model fits under plausible values of  $\beta$  suggest incorrect specification of at least one of the assumptions in M.1-M.4b.

## 5. Simulations

To evaluate the small sample performance of our estimators of  $ACE(x)$  we conducted a  $2 \times 2 \times 3 \times 3$  factorial simulation experiment, corresponding to generating data under  $\mathcal{M}_a$  or  $\mathcal{M}_b$ ;  $VE \equiv P(S(1) = 0 | S(0) = 1) \approx 0.3$  or  $0.6$ ;  $\beta = 0, 1$ , or  $3$ ;  $ACE(\cdot)$  constant and equal to  $0, 1/3$ , or  $1/2$ . Each simulation generated 1000 vectors  $W$  according to the following steps: *A)* The first 500 vectors were set at  $Z = 0$ , the second 500 were set at  $Z = 1$ . *B)*  $X$  was a single covariate generated according to the  $N(38, 6^2)$  distribution (resembling the age distribution in the VaxGen trial). *C)* Given  $X$ ,  $S(0)$  was drawn from a Bernoulli( $\theta_p(X; \mu)$ ) distribution where  $\theta_p(X; \mu)$  was as in (5) with  $\mu = (\log(1/3), 0)$  so that  $\theta_p(X; \mu)$  was constant and equal to 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). *D)*  $Y(0)$  was generated for all realizations with  $S(0) = 1$  according to the density  $f_p(y|x; \eta_0^a)$  given in (8) (under  $\mathcal{M}_a$ ) or the density  $f_p^*(y|x; \alpha, \eta_0^b)$  (under  $\mathcal{M}_b$ ) induced by  $f_p^{ai}(y|x; \eta_0^b)$  in (9) and  $w(x, y; \beta, \alpha)$  in (6). For simulations generated under  $\mathcal{M}_a$ ,  $\eta_0^a = (\lambda_0, \lambda_1, \sigma) = (2.3, 0.05, 1.0)$  (which resembles the viral load distribution for infected placebos in the VaxGen trial, where the mean and variance were 4.2 and 1.0, respectively). We detail our choices of  $\alpha$  and  $\eta_0^b$  below. *E)* Given  $X$  and  $Y(0)$ , for each realization with  $Z = 1$  and  $S(0) = 1$ ,  $S(1)$  was drawn from a Bernoulli( $w(X, Y(0); \beta, \alpha)$ ) distribution with  $w(x, y; \beta, \alpha)$  defined as in (6). In all simulations,  $\alpha_1 = \log(2)/10$ , so that for a 10 year increase in age, the odds of being in the *ai* stratum doubled. Under  $\mathcal{M}_a$ ,  $\alpha_0$  was chosen so that  $VE = 1 - \int \int w(x, y; \beta, \alpha) f_p(y|x; \eta_0^a) \varphi((x - 38)/6^2) dx dy \approx 0.3$  or  $0.6$  (where  $\varphi(\cdot)$  denotes the standard normal density). To ensure that  $VE \approx 0.3$ ,  $\alpha_0$  was set at  $-1.8, -5.8$ , or  $-13.4$ , when  $\beta$  was set at  $0, 1$ , or  $3$  respectively; and to ensure that  $VE \approx 0.6$ ,  $\alpha_0$  was set at  $-3.1, -7.4$ , or  $-16.3$ , when  $\beta = 0, 1$ , or  $3$ . For simulations generated under  $\mathcal{M}_b$ ,  $\alpha_0$  and  $\eta_0^b$  were chosen together so that  $E(Y(0)|S(0) = 1) \approx 4.2$  and  $VE \approx 0.3$  or  $0.6$ . For  $VE \approx 0.3$  we set  $(\alpha_0, \gamma_{p0}, \gamma_{p1}, \sigma_p)$  as  $(-1.8, 2.3, 0.05, 1.0), (-5.7, 2.6, 0.05, 1.0)$ , or

$(-12.1, 3.1, 0.05, 1.0)$ , when  $\beta$  was equal to 0, 1, or 3, respectively; for  $VE \approx 0.6$  we set  $(\alpha_0, \gamma_{p0}, \gamma_{p1}, \sigma_p)$  as  $(-3.1, 2.3, 0.05, 1.0)$ ,  $(-7.4, 2.9, 0.05, 1.0)$ , or  $(-17, 4.2, 0.05, 1.0)$ , when  $\beta = 0, 1, \text{ or } 3$ , respectively. *F)* For the realizations with  $Z = 1$  and  $S(1) = 1$ ,  $Y(1)$  was set equal to  $Y(0) + \Delta$ , with  $\Delta = 0, 1/3, \text{ or } 1/2$ . Note that with this choice,  $\Delta = ACE(x)$ .

MLEs were obtained using quasi-newton methods implemented in R using the function `optim()`. Under  $\mathcal{M}_a$ , when  $\beta \neq 0$ , obtaining MLEs requires maximizing over an integral that is not in closed form. Numerical integration programs written in C were used to obtain these integrals, which were then called into R. For some simulations there were local maxima. Analyses were run using multiple initial parameter values; computationally, it is important to specify good initial values. In practice we recommend starting the estimation at  $\beta = 0$ , and then to iteratively conduct estimation at increasing (or decreasing) values of  $\beta$  using as initial values the ML estimators obtained at the nearest neighbor of the current  $\beta$ . Programs can be made available by contacting the authors.

Table 1 reports Monte Carlo rejection probabilities, based on 1000 simulated datasets, of two-sided Wald tests of  $H_0 : ACE(x) = 0$  at nominal 0.05 level for the values  $x = 30, 38$ , and 55 under model  $\mathcal{M}_a$  using (5)-(8) with values of  $\beta$  set at 0, 1 or 3 (only one of them being the true value under which the data were generated) and  $\beta_v$  in (7) fixed at 0. Because of the computational time it takes to calculate the bootstrap variance of the estimate of  $E(Y(1)|S(1) = 1, X = x)$  under  $\mathcal{M}_a$ , it is not feasible to perform an extensive simulation study with M.3 correctly specified as (7) and  $\beta_v$  unknown. For this reason we conducted a small sub-study described below to evaluate the bootstrap performance, but we also evaluated the performance of standard variance estimates using the information and delta method when  $\beta_v$  was set (possibly incorrectly) to 0. Interestingly, even when the true value of  $\beta$  was different from 0 (and hence inference was conducted assuming an incorrect value of  $\beta_v$  in (7)) the type I error (i.e. the rejection probability under  $\Delta = 0$ ) was

close to the nominal 0.05 level and the tests had substantial power for detecting  $\Delta = 1/2$  at  $x = 38$  (the mean of  $X$ ). The lower power of the tests at  $x = 30$  and especially 55 is presumably because these values are far from the mean of  $X$  (with 55 being about three standard deviations away from the mean). Also, as predicted by theory, very poor inference is obtained when the values of  $\beta$  are incorrectly specified.

On a small subset of the simulated data used in Table 1 (the first 200 datasets generated with  $VE \approx 0.3$  and  $\beta = 1$ ) we examined the performance of the ML estimator of  $E(Y(1)|S(1) = 1, X = x)$  with M.3 properly specified by (7) (leaving  $\beta_v$  free) and with its variance estimated from 100 bootstrap repetitions. The coverage probabilities of Wald-based 95% confidence intervals of  $E(Y(1)|S(1) = 1, X = x)$  for  $x = 30, 38,$  and 55 were 0.950, 0.975, and 0.955, respectively.

Table 2 reports rejection probabilities for the same test under the same settings as in Table 1 except that data were generated under model  $\mathcal{M}_b$  and ML estimation was conducted under model  $\mathcal{M}_b$  using (5)-(6) and (9). Simulation results are similar, although it should be pointed out that because data were generated under different models, Tables 1 and 2 are not directly comparable. Table 3 reports rejection probabilities for  $H_0^{inter}$  and  $H_0^{global}$ , as defined in Section 4.2, based on the data and estimators used in Table 2.

## 6. 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, although interaction tests suggested that the vaccine might partially prevent



infection for non-whites and high risk subjects. Detailed study results are found in Flynn et al. (2005). Here we compare the viral load between the vaccine and placebo arms among participants (overall and within covariate subgroups) who would have been infected regardless of randomization assignment.

A total of 368 subjects were infected during the trial, and of these, 347 enrolled in the post-infection phase of the study (225 in the vaccine arm). Viral load was measured from infected participants at visits < 1 Month, 1 Month, and 2 Months post-infection diagnosis. We defined each participant's set-point viral load (the outcome of interest) as the median of all  $\log_{10}$  viral load measurements taken by the Month 2 visit and prior to initiation of antiretroviral therapy. (Results were comparable when means were used.) The viral load assay had lower and upper quantitative limits of 400 and 750,000 copies/ml, respectively. A subject's median viral load was defined exactly if the number of detectable viral load values exceeded the number undetectable. Otherwise, the median was left- or right-censored; 23 subjects had a left-censored median and 7 had a right-censored median.

There was presumably little interaction among study participants, so SUTVA was thought to be reasonable. Because this trial was randomized and double blinded, individuals' behavior and exposure to HIV were expected to be the same regardless of treatment assignment. In addition, this vaccine was designed in such a way that it could not mutate to become the virus. These facts justify A.1, although the assumption could be violated if the blinding was broken or if the vaccine induced susceptibility-enhancing immune responses for certain subjects. Some, though not all, instances in which monotonicity is violated imply that  $P(S(1) = 1) > P(S(0) = 1)$ . Those instances are detectable with power converging to 1 as the sample size goes to infinity by an  $\alpha$ -level two-sample t-test of  $H_0^* : P(S = 0|Z = 0) \geq P(S = 0|Z = 1)$ . Since the rates of infection were 241 of 3598 (6.7%) in the vaccine arm and 127 of 1805 (7.0%) in the placebo arm, a Wald test of  $H_0^*$

fails to reject. Thus, monotonicity seems reasonable.

We first estimated the unconditional average causal effect of vaccine on viral load ( $ACE$ ), defined in Section 3. Figure 1 shows the estimated  $ACE$  for  $\beta$  in  $(-3, 3)$  using A) the method of GBH with censored median viral load values set to either the lower or upper detection limit, and B) the ML estimators for the censored likelihood under  $\mathcal{M}_a$  using (5)-(8) for the relevant distributions and probabilities except not conditional on covariates. Results based on the two procedures are similar, with the method of GBH yielding slightly narrower confidence intervals presumably because censored values were truncated. The range of  $(-3, 3)$  for  $\beta$  was chosen to reflect various possibilities about the relationship between the HIV viral load distributions in the always infected and protected strata. When  $\beta$  is negative, the always infected distribution is tilted to the left of the protected distribution and the opposite happens when  $\beta$  is positive. The tilting is more marked the larger the absolute value of  $\beta$ . Absolute values of  $\beta$  as large as 3 correspond to pronounced tilting (for example, with  $\beta = 3$  the odds of being in the  $ai$  stratum versus the protected stratum for a one unit increase in viral load multiplicatively increase  $exp(3) \approx 20$ ). Thus, for example,  $\beta$  would likely be positive if individuals with relatively strong immune systems tend to have lower viral loads when infected and if the vaccine is more likely to protect these individuals from infection. On the other hand, postulating negative values of  $\beta$  would be reasonable if it is believed that the vaccine prevents infection from relatively strong/virulent viruses better than it prevents infection from weaker/a-virulent viruses. For all  $\beta$  in  $(-3, 3)$ , the null hypothesis,  $H_0: ACE = 0$ , was not rejected. The ML estimator of  $ACE$  under  $\mathcal{M}_b$  using (5)-(6) and (9) was also computed for  $\beta$  in  $(-3, 3)$ . Results using this parameterization were similar for  $\beta$  in  $(-1, 1.5)$ . However, outside this range the fitted distribution of viral load in the infected placebos under  $\mathcal{M}_b$  was bimodal, even though the observed distribution of viral loads among people infected

in the placebo arm was not bimodal. Hence, presumably for  $\beta$  outside this range, the assumed model is misspecified (as discussed in Section 4.2). The test described in HHS (which corresponds to GBH setting  $\beta = \pm\infty$ ) also did not reject  $H_0$ .

It is more compelling and interesting to estimate  $ACE(x)$  for covariate levels under which vaccine efficacy,  $P(S(1) = 0 | S(0) = 1, X = x)$ , is large; it is hypothesized that vaccine-induced antibodies that partially protect against infection also may have a beneficial effect to lower viral load. Among non-whites vaccine efficacy was estimated as 0.487. The vaccine also appeared more effective for those with higher baseline risk scores (the number of self-reported high infection-risk behavior activities, taking integer values from 0-9). There is an hypothesized biological mechanism for why the vaccine’s ability to prevent infection and lower viral load might vary by risk behavior: natural exposure to HIV may “prime” the immune system, which is “boosted” by the vaccine to provide extra protection (Rowland-Jones et al., 1998). Hence, it is of interest to apply our methods to estimate the  $ACE(x)$  at different risk levels among non-whites.

The first row of Figure 2 contains sensitivity analyses of the estimated  $ACE(x)$  at different risk scores based on the non-white cohort data only, under distributional assumptions  $\mathcal{M}_a$  using (5)-(8) with  $x$  defined as risk score. For non-whites with risk scores of 2, if  $\beta < -1/2$  then there is evidence that  $ACE(x) > 0$ . The second row of Figure 2 is similar to the first, only this analysis was performed under  $\mathcal{M}_b$  using (5)-(6) and (9) for the relevant distributions and probabilities. Under  $\mathcal{M}_b$ , if  $\beta \geq 1$ , there is evidence that the vaccine causes lower viral loads among non-whites with risk scores of 2, 3, and 4. In addition, based on the analysis under  $\mathcal{M}_b$ , if  $\beta \geq 1.5$ ,  $H_0^{global}$  is rejected at the 0.05 level. There is insufficient evidence, however, to conclude that  $ACE(x)$  varies by risk score; under  $\mathcal{M}_b$ , at all values of  $\beta$  in  $(-3, 3)$ , we fail to reject  $H_0^{inter}$  (P-values  $> 0.18$ ).

The discrepancies between the first and second rows of Figure 2 are primarily due to

different model choices for  $f_v(y|x; \eta_1)$ . In the first row under  $\mathcal{M}_a$  using (7),  $f_v(y|x; \eta_1)$  is modeled with six parameters, leading to a much more flexible estimate of  $E(Y(1)|S(1) = 1, X = x)$  than under  $\mathcal{M}_b$  using (9) (see Figure 3). When we perform the sensitivity analyses under  $\mathcal{M}_a$  except setting  $\beta_v$  in (7) equal to 0 (as in Table 1 of our simulations, which is equivalent to modeling  $f_v(y|x; \eta_1)$  with the normal linear model (9)), then the estimated  $ACE(x)$  is very similar to the estimate under  $\mathcal{M}_b$  (see Figure 2, row 3).

It should be noted that for large negative values of  $\beta$ , our parameterization choices under  $\mathcal{M}_b$  are probably inadequate. The sharp shifts in the  $ACE(x)$  (seen in the plot with Risk=4) for  $\beta < -1.5$  are presumably due to either M.2 or M.4b being misspecified.

## 7. Discussion

In this paper we have considered estimation of  $ACE(x)$  under parametric models for the counterfactuals. We have motivated our methods with an application in which  $Y$  is continuous. Our methods apply also to discrete outcomes, provided adequate parametric models are used in M.3 and M.4. As we have seen, there are some inherent challenges to using parametric methods for estimating  $ACE(x)$ . Under  $\mathcal{M}_a$ , there are identification and over-fitting issues. Under  $\mathcal{M}_b$ , it may be difficult to find models that are compatible with the observed data for the complete range of  $\beta$ . One could imagine a third parametric model, say  $\mathcal{M}_c$ , based on assumptions (1), (2), and A.1, modeling  $f_v(\cdot)$ ,  $f_p^{ai}(\cdot)$ , and instead of  $w(\cdot)$ , modeling  $f_p^{prot}(\cdot)$  (the distribution of viral loads for placebos in the protected principal stratum), with a sensitivity parameter specifying some discrepancy between  $f_p^{prot}(\cdot)$  and  $f_p^{ai}(\cdot)$ . This is very similar to model  $\mathcal{M}_b$  (where we specified  $f_p^{ai}(\cdot)$  and  $w(\cdot)$ , inducing  $f_p^{prot}(\cdot)$ ). Although under this model there is no longer the requirement that  $w(\cdot) > 0$  for all  $x$  and  $y$ , care would still have to be taken to make distributional assumptions compatible with the observed data for the entire range of  $\beta$  (similar to model

$\mathcal{M}_b$ ). Instead of considering inference using distributional shape assumptions on viral loads we could consider estimation under less restrictive semiparametric models that only make assumptions about the conditional means of the counterfactuals given covariates in the  $ai$  principal stratum. One advantage of the semiparametric approach is that it gives inferences (in particular, tests of the null hypothesis  $H_0 : ACE(x)$  for all  $x$ ), that are valid under less stringent assumptions about the distribution of the outcome in the  $ai$  stratum. In future work we will describe semiparametric estimation of  $ACE(x)$ .

Both the semiparametric and the parametric methods require assuming that  $\beta$  in A.2 is known. In its fullest sense, a sensitivity analysis includes not just varying  $\beta$ , but also varying the form of  $w(x, y; \beta)$ . To be useful,  $w(x, y; \beta)$  should be chosen so that the sensitivity parameter  $\beta$  has a meaningful interpretation, allowing one to intuitively choose a range over which to perform the analyses; with  $w(x, y; \beta)$  defined by (6),  $\beta$  has a log-odds ratio interpretation.

These sensitivity analyses can be thought of as examining departures from assumption A.3, that  $Y(0) \perp\!\!\!\perp S(1) | S(0) = 1, X$ . If A.3 holds, then one can simply perform standard regression analyses on the infected subjects. Adding to  $X$  additional covariates thought to be associated with both vaccine efficacy and viral load may make A.3 more believable. However, this does not imply that adding covariates to the analysis restricts the range of values of  $\beta$  considered plausible (Scharfstein et al., 1999). If monotonicity is not reasonable, then in order to identify  $ACE(x)$  one will need to make other assumptions about the joint distribution of  $(S(0), S(1))$  given  $X$ . One such assumption is explainable non-random infection (Robins, 1998; Hayden et al., 2005). In the absence of covariates and monotonicity, sharp non-parametric bounds have been derived by Jemai (unpublished Ph.D. dissertation) and Zhang and Rubin (2003). Jemai’s dissertation also discusses methods for conducting sensitivity analyses.

Other possible applications of these methods include: mother-to-child HIV transmission studies, comparing outcomes in always surviving infants; antiretroviral clinical trials, comparing metrics of HIV resistance mutations in subjects always failing treatment regimens; and cancer research, comparing disease severity between a preventative treatment and placebo among those who would have developed cancer regardless of randomization. In general, these methods are applicable to intervention studies with post-randomization selection criteria where SUTVA and monotonicity are thought to hold.

#### ACKNOWLEDGEMENTS

The authors thank Michael Hudgens and the referees for helpful discussions. This work was supported by NIH grants 1 RO1 AI054165-01 (Gilbert), R01 AI32475 (Rotnitzky), and R29 GM48704 (Jemai and Rotnitzky). This work was partially completed while Andrea Rotnitzky was visiting the Department of Economics at Di Tella University. She would like to thank the authorities of the school for providing a stimulating environment.

#### REFERENCES

- Frangakis, C.E. and Rubin, D.B. (2002). Principal stratification in causal inference. *Biometrics* **58**, 21–29.
- Flynn, N.M., Forthal, C.D., Harro, F.N., et al. (2005). Placebo-controlled trial of a phase 3 trial of a recombinant glycoprotein 120 vaccine to prevent HIV-1 infection. *Journal of Infectious Diseases* **191**, 647–649.
- 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.
- Hayden, D., Pauler, D.K., and Schoenfeld, D. (2005). An estimator for treatment comparisons among survivors in randomized trials. *Biometrics* **61**, 305–310.
- 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.
- O'Brien, T.R., Blattner, W.A., Waters, D., et al. (1996). Serum HIV-1 RNA levels and time to development of AIDS in the Multicenter Hemophilia Cohort Study. *Journal of the American Medical Association* **276**, 105–110.
- Quinn, T.C., Wawer, M.J., Sewankambo, N., et al. (2002). Viral load and heterosexual transmission of human immunodeficiency virus type 1. *New England Journal of Medicine* **342**, 921–929.
- Robins, J.M. (1986). A new approach to causal inference in mortality studies with sus-

- tained 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. (1998). Correction for non-compliance in equivalence trials. *Statistics in Medicine* **17**, 269–302.
- 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.
- Rowland-Jones, S.L., Dong, T., Fowke, K.R., et al. (1998). Cytotoxic T cell responses to multiple conserved HIV epitopes in HIV-resistant prostitutes in Nairobi. *Journal of Clinical Investigation* **102**, 1758–1765.
- Rubin, D.B. (1978). Bayesian inference for causal effects: the role of randomization. *The Annals of Statistics* **6**, 34–58.
- Rubin, D.B. (1986). Statistics and causal inference: which ifs have causal answers. *Journal of the American Statistical Association* **81**, 961–962.
- 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., and Robins, J.M. (1999). Adjusting for nonignorable drop-out using semiparametric nonresponse models. *Journal of the American Statistical Association* **94**, 1096–1146.

Zhang, J.L. and 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.

## APPENDIX

Under  $\mathcal{M}_a$ ,

$$\begin{aligned}
& f_{Y|S,Z,X}(y|S=1, Z=0, X=x) \\
&= f_{Y(0)|S(0)=1,Z,X}(y|S(0)=1, Z=0, X=x) \quad \text{by (1)} \\
&= f_{Y(0)|S(0)=1,X}(y|S(0)=1, X=x) \quad \text{by (2)} \\
&= f_p(y|x; \eta_0^a) \quad \text{by M.4a.}
\end{aligned}$$

Under  $\mathcal{M}_b$ ,

$$\begin{aligned}
& f_{Y|S,Z,X}(y|S=1, Z=0, X=x) \\
&= f_{Y(0)|S(0)=1,Z,X}(y|S(0)=1, Z=0, X=x) \quad \text{by (1)} \\
&= f_{Y(0)|S(0)=1,X}(y|S(0)=1, X=x) \quad \text{by (2)} \\
&= \frac{w^{-1}(x, y; \beta, \alpha)w(x, y; \beta, \alpha)f_{Y(0)|S(0)=1,X}(y|S(0)=1, X=x)}{\int w(x, y; \beta, \alpha)f_{Y(0)|S(0)=1,X}(y|S(0)=1, X=x)dy} \Bigg/ \\
&\quad \frac{\int w^{-1}(x, y; \beta, \alpha)w(x, y; \beta, \alpha)f_{Y(0)|S(0)=1,X}(y|S(0)=1, X=x)dy}{\int w(x, y; \beta, \alpha)f_{Y(0)|S(0)=1,X}(y|S(0)=1, X=x)dy} \quad \text{if } w(\cdot) > 0 \\
&= \frac{w^{-1}(x, y; \beta, \alpha)f_{Y(0)|S(0)=S(1)=1,X}(y|S(0)=S(1)=1, X=x)}{\int w^{-1}(x, y; \beta, \alpha)f_{Y(0)|S(0)=S(1)=1,X}(y|S(0)=S(1)=1, X=x)dy} \quad \text{by A.2, M.2} \\
&= \frac{w^{-1}(x, y; \beta, \alpha)f_p^{ai}(y|x; \eta_0^b)}{\int w^{-1}(x, y; \beta, \alpha)f_p^{ai}(y|x; \eta_0^b)dy} \quad \text{by M.4b} \\
&\equiv f_p^*(y|x; \alpha, \eta_0^b).
\end{aligned}$$

Under  $\mathcal{M}_a$  or  $\mathcal{M}_b$ ,

$$\begin{aligned}
& f_{Y|S,Z,X}(y|S=1, Z=1, X=x) \\
&= f_{Y(1)|S(1)=1,Z,X}(y|S(1)=1, Z=1, X=x) && \text{by (1)} \\
&= f_{Y(1)|S(1)=1,X}(y|S(1)=1, X=x) && \text{by (2)} \\
&= f_v(y|x; \eta_1) && \text{by M.3.}
\end{aligned}$$

$$\begin{aligned}
& P_{S|Z,X}(S=1|Z=0, X=x) \\
&= P_{S(0)|Z,X}(S(0)=1|Z=0, X=x) && \text{by (1)} \\
&= P_{S(0)|Z,X}(S(0)=1|X=x) && \text{by (2)} \\
&= \theta_p(x; \mu) && \text{by M.1.}
\end{aligned}$$

$$\begin{aligned}
& P_{S|Z,X}(S=1|Z=1, X=x) \\
&= P_{S(1)|Z,X}(S(1)=1|Z=1, X=x) && \text{by (1)} \\
&= P_{S(1)|Z,X}(S(1)=1|X=x) && \text{by (2)} \\
&= P_{S(0)|X}(S(0)=1|X=x) P_{S(1)|S(0),X}(S(1)=1|S(0)=1, X=x) && \text{by A.1} \\
&= \theta_p(x; \mu) \times \int P_{S(1)|S(0),Y(0),X}(S(1)=1|S(0)=1, Y(0)=y, X=x) \\
&\quad f_{Y(0)|S(0)=1,X}(y|S(0)=1, X=x) dy && \text{by M.1} \\
&= \begin{cases} \theta_p(x; \mu) \int w(x, y; \beta, \alpha) f_p(y|x; \eta_0^a) dy && \text{by A.2, M.2, M.4a} \\ \theta_p(x; \mu) \int w(x, y; \beta, \alpha) f_p^*(y|x; \alpha, \eta_0^b) dy && \text{by A.2, M.2, M.4b} \end{cases}
\end{aligned}$$

Table 1: Size/power for detecting a 0, 1/3, and 1/2 mean shift alternative of the Wald test of  $H_0 : ACE(x) = 0$  under  $\mathcal{M}_a$  with  $\beta_v$  in (7) set equal to 0.

VE	True	Presumed	$\Delta=0$			$\Delta=1/3$			$\Delta=1/2$			
	$\beta$	$\beta$	$x = 30$	$x = 38$	$x = 55$	$x = 30$	$x = 38$	$x = 55$	$x = 30$	$x = 38$	$x = 55$	
$\sim 0.3$	<b>0</b>	<b>0</b>	<b>0.047</b>	<b>0.052</b>	<b>0.050</b>	<b>0.251</b>	<b>0.634</b>	<b>0.129</b>	<b>0.506</b>	<b>0.936</b>	<b>0.230</b>	
	0	1	0.260	0.342	0.074	0.053	0.098	0.073	0.096	0.360	0.145	
	0	3	0.520	0.588	0.086	0.137	0.097	0.065	0.066	0.080	0.106	
	1	0	0.383	0.479	0.048	0.868	0.994	0.115	0.956	1	0.237	
	<b>1</b>	<b>1</b>	<b>0.056</b>	<b>0.056</b>	<b>0.059</b>	<b>0.264</b>	<b>0.605</b>	<b>0.086</b>	<b>0.477</b>	<b>0.919</b>	<b>0.159</b>	
	1	3	0.146	0.132	0.060	0.066	0.173	0.075	0.123	0.460	0.128	
	3	0	0.832	0.951	0.067	0.993	1	0.161	0.998	1	0.288	
	3	1	0.178	0.275	0.074	0.653	0.949	0.114	0.838	0.998	0.197	
	<b>3</b>	<b>3</b>	<b>0.048</b>	<b>0.073</b>	<b>0.083</b>	<b>0.184</b>	<b>0.521</b>	<b>0.089</b>	<b>0.379</b>	<b>0.860</b>	<b>0.161</b>	
	$\sim 0.6$	<b>0</b>	<b>0</b>	<b>0.058</b>	<b>0.042</b>	<b>0.058</b>	<b>0.193</b>	<b>0.462</b>	<b>0.128</b>	<b>0.350</b>	<b>0.807</b>	<b>0.195</b>
		0	1	0.462	0.752	0.085	0.144	0.122	0.061	0.073	0.049	0.086
		0	3	0.825	0.979	0.141	0.507	0.682	0.061	0.318	0.336	0.059
1		0	0.483	0.813	0.062	0.828	0.997	0.193	0.921	1	0.303	
<b>1</b>		<b>1</b>	<b>0.058</b>	<b>0.071</b>	<b>0.043</b>	<b>0.213</b>	<b>0.471</b>	<b>0.107</b>	<b>0.357</b>	<b>0.775</b>	<b>0.167</b>	
1		3	0.217	0.391	0.051	0.057	0.109	0.171	0.102	0.060	0.096	
3		0	0.987	1	0.114	1	1	0.335	1	1	0.487	
3		1	0.433	0.635	0.068	0.999	1	0.334	1	1	0.485	
<b>3</b>		<b>3</b>	<b>0.049</b>	<b>0.053</b>	<b>0.067</b>	<b>0.190</b>	<b>0.492</b>	<b>0.115</b>	<b>0.339</b>	<b>0.807</b>	<b>0.189</b>	

Table 2: Size/power for detecting a 0, 1/3, and 1/2 mean shift alternative of the Wald test of  $H_0 : ACE(x) = 0$  under  $\mathcal{M}_b$ .

VE	True	Presumed	$\Delta=0$			$\Delta=1/3$			$\Delta=1/2$			
	$\beta$	$\beta$	$x = 30$	$x = 38$	$x = 55$	$x = 30$	$x = 38$	$x = 55$	$x = 30$	$x = 38$	$x = 55$	
$\sim 0.3$	<b>0</b>	<b>0</b>	<b>0.058</b>	<b>0.055</b>	<b>0.047</b>	<b>0.308</b>	<b>0.673</b>	<b>0.124</b>	<b>0.559</b>	<b>0.949</b>	<b>0.211</b>	
	0	1	0.232	0.363	0.051	0.054	0.084	0.096	0.110	0.350	0.157	
	0	3	0.492	0.710	0.153	0.253	0.323	0.133	0.213	0.272	0.159	
	1	0	0.471	0.573	0.048	0.881	0.988	0.095	0.956	1	0.172	
	<b>1</b>	<b>1</b>	<b>0.056</b>	<b>0.061</b>	<b>0.050</b>	<b>0.269</b>	<b>0.547</b>	<b>0.121</b>	<b>0.449</b>	<b>0.853</b>	<b>0.197</b>	
	1	3	0.314	0.432	0.096	0.384	0.439	0.140	0.441	0.579	0.200	
	3	0	1	1	0.391	1	1	0.173	1	1	0.092	
	3	1	0.422	0.422	0.051	0.767	0.945	0.032	0.888	0.994	0.057	
	<b>3</b>	<b>3</b>	<b>0.056</b>	<b>0.052</b>	<b>0.058</b>	<b>0.186</b>	<b>0.502</b>	<b>0.121</b>	<b>0.351</b>	<b>0.831</b>	<b>0.201</b>	
	$\sim 0.6$	<b>0</b>	<b>0</b>	<b>0.057</b>	<b>0.058</b>	<b>0.069</b>	<b>0.209</b>	<b>0.507</b>	<b>0.143</b>	<b>0.387</b>	<b>0.824</b>	<b>0.212</b>
		0	1	0.466	0.742	0.117	0.139	0.136	0.066	0.063	0.050	0.098
		0	3	0.797	0.988	0.210	0.492	0.697	0.127	0.326	0.382	0.120
1		0	0.644	0.900	0.083	0.896	0.998	0.174	0.964	1	0.257	
<b>1</b>		<b>1</b>	<b>0.061</b>	<b>0.049</b>	<b>0.062</b>	<b>0.180</b>	<b>0.388</b>	<b>0.113</b>	<b>0.307</b>	<b>0.686</b>	<b>0.167</b>	
1		3	0.284	0.460	0.129	0.132	0.110	0.121	0.124	0.170	0.164	
3		0	1	1	0.062	1	1	0.048	1	1	0.070	
3		1	0.505	0.655	0.038	0.750	0.935	0.076	0.838	0.982	0.116	
<b>3</b>		<b>3</b>	<b>0.063</b>	<b>0.064</b>	<b>0.056</b>	<b>0.145</b>	<b>0.301</b>	<b>0.105</b>	<b>0.240</b>	<b>0.545</b>	<b>0.168</b>	

Table 3: Size of the Wald test for  $ACE(x)$  independent of  $x$  ( $H_0^{inter}$ :  $\gamma_{v1} = \gamma_{p1}$ ), and size/power of the likelihood ratio test of  $ACE(x) = 0$  for all  $x$  ( $H_0^{global}$ :  $\gamma_v = \gamma_p$ ), under  $\mathcal{M}_b$ .

VE	True	Presumed	$H_0^{inter}$	$H_0^{global}$		
	$\beta$	$\beta$		$\Delta = 0$	$\Delta = 1/3$	$\Delta = 1/2$
$\sim 0.3$	<b>0</b>	<b>0</b>	<b>0.063</b>	<b>0.051</b>	<b>0.561</b>	<b>0.907</b>
	0	1	0.068	0.283	0.093	0.289
	0	3	0.128	0.569	0.460	0.530
	1	0	0.162	0.491	0.983	1
	<b>1</b>	<b>1</b>	<b>0.057</b>	<b>0.052</b>	<b>0.456</b>	<b>0.780</b>
	1	3	0.135	0.261	0.311	0.453
	3	0	0.948	1	1	1
	3	1	0.187	0.307	0.874	0.979
	<b>3</b>	<b>3</b>	<b>0.055</b>	<b>0.057</b>	<b>0.420</b>	<b>0.769</b>
	$\sim 0.6$	<b>0</b>	<b>0</b>	<b>0.075</b>	<b>0.056</b>	<b>0.430</b>
0		1	0.087	0.662	0.124	0.076
0		3	0.139	0.975	0.635	0.368
1		0	0.146	0.824	0.994	1
<b>1</b>		<b>1</b>	<b>0.051</b>	<b>0.059</b>	<b>0.317</b>	<b>0.599</b>
1		3	0.118	0.383	0.104	0.151
3		0	0.878	1	1	1
3		1	0.170	0.507	0.901	0.971
<b>3</b>		<b>3</b>	<b>0.057</b>	<b>0.056</b>	<b>0.275</b>	<b>0.528</b>

**Figure 1.** Sensitivity analysis estimates and 95% confidence intervals of the  $ACE$  for the complete VaxGen cohort using both the method of GBH and parameterizing with  $\mathcal{M}_a$ .

**Figure 2.** Sensitivity analyses of the  $ACE(x)$  at different risk behavior levels in the non-white cohort. The first row is under distributional assumptions  $\mathcal{M}_a$  using (5)-(8) for the relevant distributions and probabilities. The second row is under  $\mathcal{M}_b$  using (5)-(6) and (9). The third row is under  $\mathcal{M}_a$  using (5)-(8) except fixing  $\beta_v = 0$  in (7).

**Figure 3.** Estimates for the expected viral load among infected non-whites in the vaccine arm conditional on risk score,  $E(Y(1)|S(1) = 1, X = x)$ , based on model (7) (solid line) and model (9) (dashed line). The circles are the observed viral loads (the three viral load values below 2.5 were censored).

Figure 1.

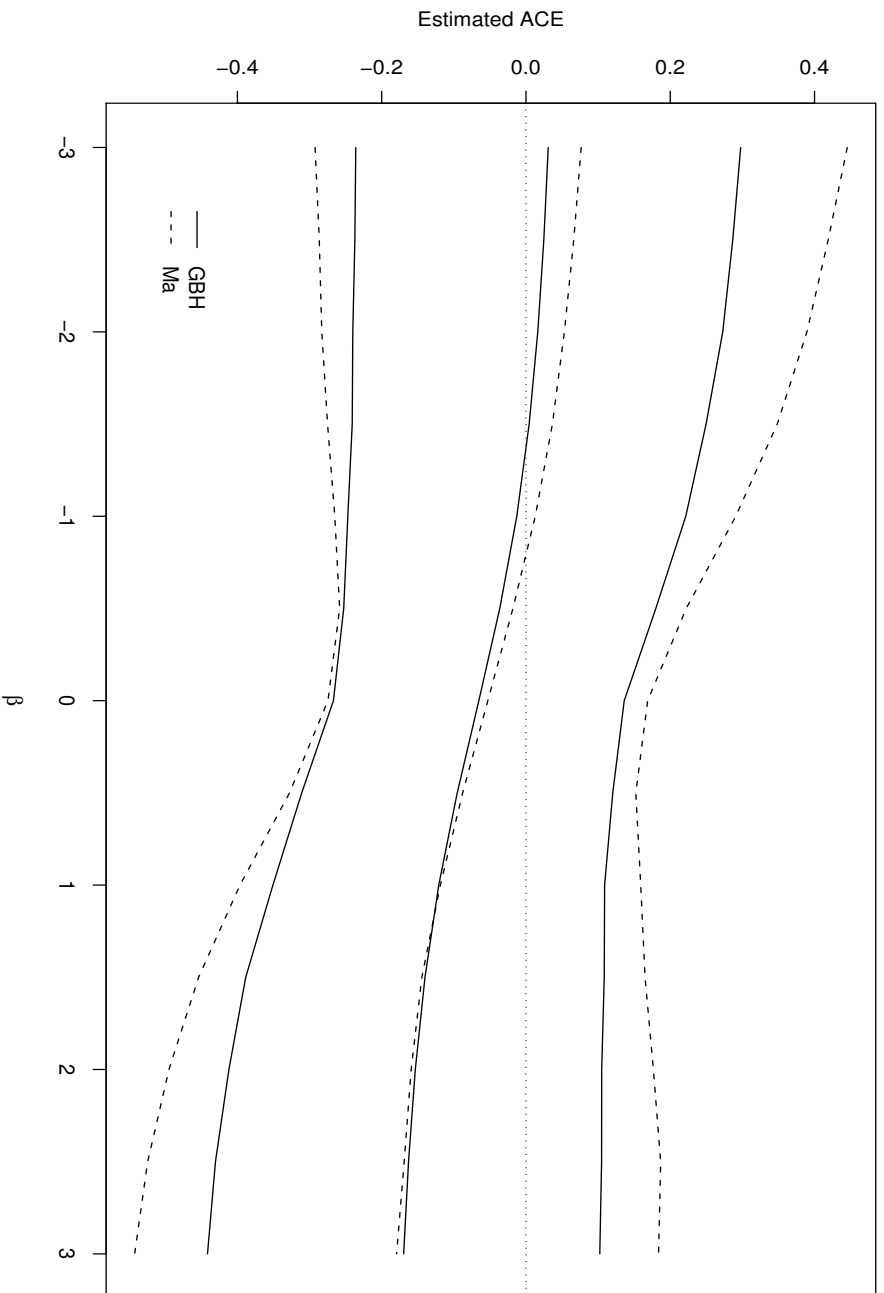


Figure 2.

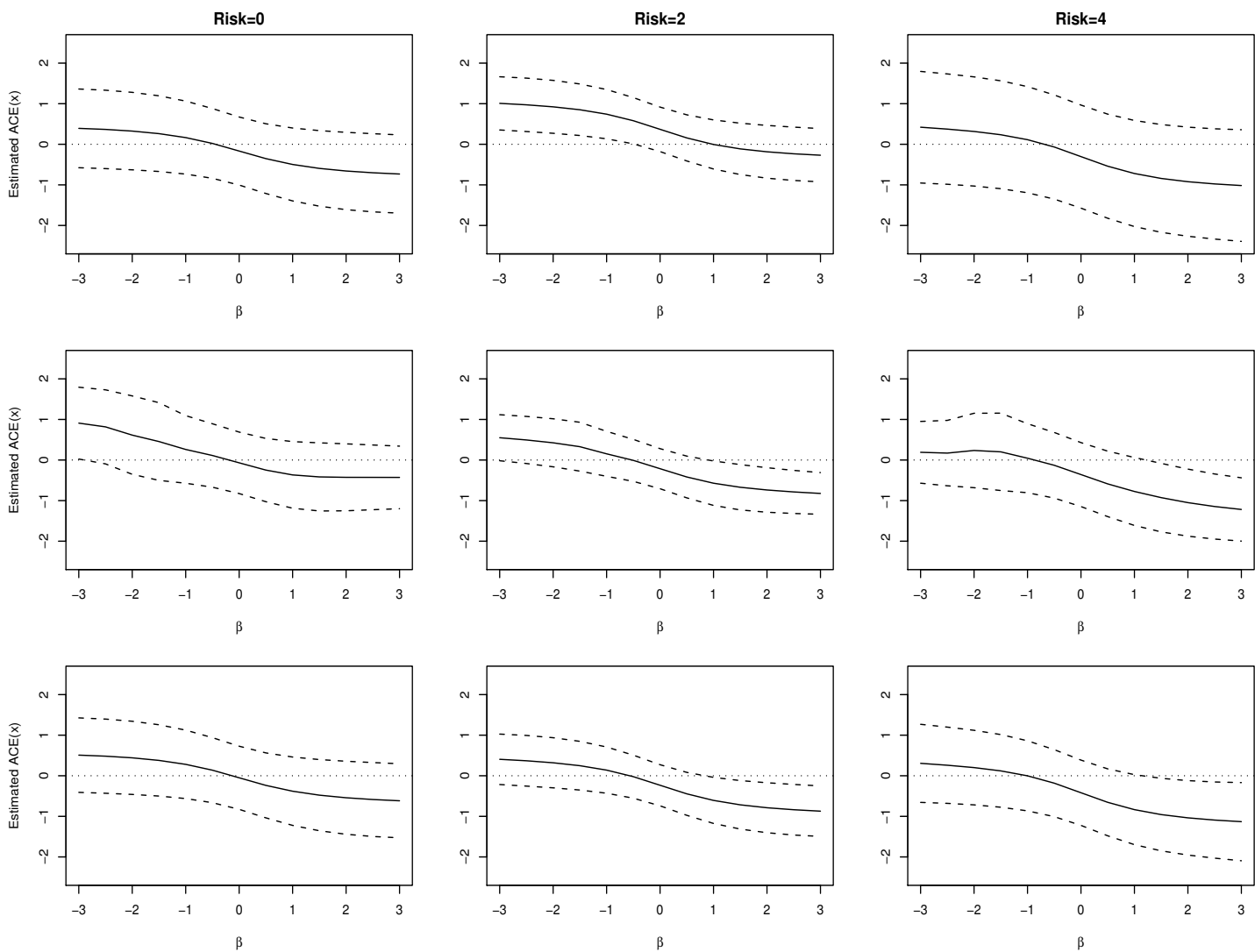




Figure 3.

