Biometrics 59, 531–541
September 2003

Sensitivity Analysis for the Assessment of Causal Vaccine Effects on Viral Load in HIV Vaccine Trials

Peter B. Gilbert,[1,*] Ronald J. Bosch,[2] and Michael G. Hudgens[1]

[1]Fred Hutchinson Cancer Research Center and Department of Biostatistics, University of Washington, Seattle, Washington, 98109, U.S.A.
[2]Center for Biostatistics in AIDS Research and Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts 02115, U.S.A.
[*]*email:* pgilbert@scharp.org

Summary. Vaccines with limited ability to prevent HIV infection may positively impact the HIV/AIDS pandemic by preventing secondary transmission and disease in vaccine recipients who become infected. To evaluate the impact of vaccination on secondary transmission and disease, efficacy trials assess vaccine effects on HIV viral load and other surrogate endpoints measured after infection. A standard test that compares the distribution of viral load between the infected subgroups of vaccine and placebo recipients does not assess a causal effect of vaccine, because the comparison groups are selected after randomization. To address this problem, we formulate clinically relevant causal estimands using the principal stratification framework developed by Frangakis and Rubin (2002, *Biometrics* **58,** 21–29), and propose a class of logistic selection bias models whose members identify the estimands. Given a selection model in the class, procedures are developed for testing and estimation of the causal effect of vaccination on viral load in the principal stratum of subjects who would be infected regardless of randomization assignment. We show how the procedures can be used for a sensitivity analysis that quantifies how the causal effect of vaccination varies with the presumed magnitude of selection bias.

Key words: Causal inference; Posttreatment selection bias; Principal stratification.

1. Introduction

Consider a randomized, two-arm, placebo-controlled clinical trial to evaluate efficacy of a preventive HIV vaccine. The first two trials of this kind began in 1998 and 1999, and are ongoing (Francis et al., 1998). For each trial, the primary objective is to assess the vaccine's impact on the incidence of HIV infection (Rida and Lawrence, 1995). Another objective of these trials and future trials is to assess the vaccine's impact on viral load after acquisition of HIV (Nabel, 2001); viral load is the concentration of HIV in blood or another body compartment. This objective is important because natural history studies have shown that the viral load of an infected person predicts infectiousness (Quinn et al., 2000) and the rate of disease progression (Mellors et al., 1997), and several animal studies have identified vaccines that failed to prevent infection but successfully controlled viremia and prevented disease (Shiver et al., 2002). Therefore, a vaccine effect to lower viral load may be beneficial, whereas an effect to increase viral load may hasten or exacerbate disease. The risk of harmful vaccine "enhancement" of viral load is genuine (Burke, 1992), and has been observed for several viral vaccines (Mascola et al., 1993).

The impact of vaccination on viral load can be studied in several ways. The data available for analysis are longitudinal quantitative measurements of viral load in subjects who become infected. Two main inferential approaches are intent-to-treat (ITT) analyses of all randomized subjects and conditional analyses of infected subjects only. The ITT approach assesses the causal effect of randomizing to vaccine. However, the majority (likely $> 80\%$) of randomized subjects will have zero viral load because they do not become infected during the trial, which can give ITT analyses low power for detecting certain alternatives of interest. Also, in ITT analyses, two very different populations (uninfected and infected subjects) are placed on the same response scale. Consequently, the ITT analysis assumes that the outcomes of both absence of infection (with zero viral load) and infection with viral load below the quantification limit of the assay are approximately equally prognostic for disease progression. This assumption is difficult to justify, because the initial vaccine-induced suppression of viral load may be lost due to HIV evolution (Barouch et al., 2002). Alternatively, rank-based ITT methods could be used that assign the lowest two ranks to absence of infection and viral load below the assay limit. However, to achieve greater power, and to study the causal effect of randomizing to vaccine in a subpopulation of persons who would become infected, as described below, in this article we consider conditional analyses.

Conditioning on infection poses a major challenge to making an unbiased inference of the vaccine effect on viral load, because the analyzed groups are selected by the post-randomization event HIV infection. This problem of post-treatment selection bias is common in biomedical studies (Rosenbaum, 1984; Robins and Greenland, 1992), and implies that a comparison of viral load between infected subgroups, which measures the "net vaccine effect," does not have a causal interpretation. In particular, partial efficacy of the vaccine to prevent HIV infection can bias the comparison. For example, the vaccine may prevent infections in individuals with strong immune systems, but allow infections in individuals with relatively weak immune systems. If a weaker immune system correlates with a higher viral load upon infection, then the viral loads in infected subjects will be selectively shifted upwards in vaccine relative to placebo recipients. On the other hand, selection bias could occur in the opposite direction, for example if the vaccine better protects against virulent than mild viruses. Therefore, a two-sample test comparing viral loads between infected groups may give a misleading impression that vaccination enhances or suppresses viral burden. Or, the test may fail to detect a meaningful vaccine effect.

Frangakis and Rubin (FR) (2002) developed a framework for causal inference that can be used for studying a causal effect of vaccine on viral load that adjusts for the post-randomization selection bias. This framework defines causal estimands using potential outcomes (Rubin, 1974, 1978; Holland, 1986). For the present problem, each trial participant has a potential infection status under each randomization assignment. Additionally, subjects who would be infected under randomization to vaccine have a potential viral load under vaccine assignment, and subjects who would be infected under randomization to placebo have a potential viral load under placebo assignment. Within FR's framework, a *causal vaccine effect* on viral load is defined as a comparison of potential viral loads under the two randomization assignments for a subgroup of subjects with a common pair of potential infection status outcomes; FR referred to such a group as a *principal stratum*.

Hudgens, Hoering, and Self (HHS) (2003, in press) developed tests for a causal vaccine effect on viral load in the "always-infected" principal stratum of subjects who would be infected regardless of randomization to vaccine or placebo. Under plausible assumptions described in Section 2, vaccine recipients who become HIV infected would also be infected had they received placebo. Consequently, inferences drawn for the always-infected subpopulation address a practical question for individuals vaccinated in a public health program: "If I acquire HIV despite vaccination, what is the viral load compared to if I had forgone vaccination?" We consider inference on causal estimands defined for the always-infected principal stratum, which are defined in terms of potential outcomes in Section 2.

The causal estimands are not identified, because membership of an infected placebo recipient in the always-infected principal stratum is unknown (i.e., the infection status had the subject been randomized to vaccine is unknown). This problem can be addressed by modeling the probability that an infected placebo recipient is in the always-infected stratum as a function of the potential viral load under randomization to placebo. HHS implicitly took this approach, by defining two selection models that express bounds for the maximum plausible levels of selection bias. Under these models, which identify the estimands, HHS developed testing procedures for assessing differences in the potential viral load distributions of always-infected subjects under the two randomization assignments.

Testing the null hypothesis presuming the maximal degree of selection bias is practically very useful, because rejection implies a significant effect of vaccination above and beyond any plausible selective effects. However, the actual degree of bias is likely less than that specified by an extreme model, so that HHS's tests may sacrifice power. Achieving maximal power is especially important for key subgroup analyses, such as by gender (Sterling et al., 2001), route of exposure, or host genotype, and for analyses of seminal viral load, given the higher variability of seminal versus plasma viral load (Coombs et al., 1998). Therefore, it is important to also consider selection models that reflect intermediate degrees of selection bias, which may be more realistic and will allow for more powerful statistical tests. In this article, we develop a method for sensitivity analysis that considers a continuous range of possible selective effects spanning from no bias to maximal plausible bias, as considered by HHS. In different contexts, Rosenbaum and Rubin (1983), Scharfstein, Rotnitzky, and Robins (1999), and Goetghebeur et al. (2000) also developed methods of continuously indexed sensitivity analysis of inferences on causal effects.

The article is organized as follows. Section 2 defines causal estimands and shows that they are identified from three assumptions and a biased sampling model that specifies the nature and degree of selection bias. A class of logistic-biased sampling models is described, which is indexed by an interpretable sensitivity parameter β that can be chosen to represent any magnitude of selection bias ranging between extreme positive and negative bias. Given a particular model in the class, Section 3 describes procedures for testing the corresponding null hypothesis of no causal effect of vaccination on viral load. A plot of the test statistic (or p-value) versus β provides a sensitivity analysis to help discern if the data support a causal effect that is robust to plausible postrandomization selective effects. Section 3 also shows how the sensitivity analysis can be based on estimation rather than testing. Section 4 evaluates the proposed testing procedures in simulations, and Section 5 illustrates a sensitivity analysis on a simulated vaccine trial dataset. Simulated data are used by necessity, since real data will not be available until the ongoing trials complete.

2. Causal Estimands and Biased Sampling Models for Sensitivity Analysis

In this section, we first use the framework of potential outcomes to formulate the causal estimands, which are defined in terms of the distributions of potential viral loads (PVLs) given assignment to receive vaccine or placebo for subjects in the always-infected principal stratum. We then delineate assumptions that identify the estimands. Identifiability is achieved by equating the PVL distribution if assigned to recieve vaccine to the viral load distribution in infected vaccine recipients, and by specifying the PVL distribution if assigned to receive

placebo as a known weighted function of the viral load distribution in infected placebo recipients. These models use the fact that the trial is randomized and blinded, and assume that assignment to vaccine does not increase the probability of HIV infection.

2.1. *Definition and Identifiability of Causal Estimands*

First we define the potential outcomes of the trial participants. Let $\mathbf{Z}$ be the vector of vaccination assignments for the N randomized subjects, with ith element Z_i ($Z_i = v$, vaccine; $Z_i = p$, placebo). Let $\mathbf{S}(\mathbf{Z})$ be the N-vector with ith element $S_i(\mathbf{Z})$, which is the indicator of whether the ith subject would be infected given $\mathbf{Z}$. For subjects with $S_i(\mathbf{Z}) = 1$, let $Y_i(\mathbf{Z}, \mathbf{S})$ be the PVL given $\mathbf{Z}$ and $\mathbf{S} = \mathbf{S}(\mathbf{Z})$. In order to limit the possible potential outcomes for each subject, we adopt Rubin's (1978) stable unit treatment value assumption (SUTVA) throughout. It states that $S_i(\mathbf{Z}) = S_i(\mathbf{Z}')$ whenever $Z_i = Z_i'$, and $Y_i(\mathbf{Z}, \mathbf{S}) = Y_i(\mathbf{Z}', \mathbf{S}')$ whenever $Z_i = Z_i'$ and $S_i(Z_i) = S_i'(Z_i) = 1$. SUTVA implies that potential outcomes for each subject i are unrelated to the assignment Z_j of other subjects, and allows $S_i(\mathbf{Z})$ and $Y_i(\mathbf{Z}, \mathbf{S})$ to be written as $S_i(Z_i)$ and $Y_i(Z_i)$, respectively. Therefore, under SUTVA, each subject has two potential infection outcomes ($S_i(v)$, $S_i(p)$), and at most two PVL outcomes ($Y_i(v)$, $Y_i(p)$). For each subject only one of $S_i(v)$ or $S_i(p)$ is observed, denoted $S_i^{\text{obs}} \equiv S_i(Z_i)$; in the subgroup with $S_i^{\text{obs}} = 1$, $Y_i^{\text{obs}} \equiv Y_i(Z_i)$ is observed. Note that $Y_i(v)\{Y_i(p)\}$ is defined only if $S_i(v) = 1\{S_i(p) = 1\}$.

By Property 2 of FR, a comparison between the ordered sets $\{Y_i(v) : S_i(v) = S_i(p) = 1\}$ and $\{Y_i(p) : S_i(v) = S_i(p) = 1\}$ is a causal effect, because it is made within a principal stratum. For subjects in the always-infected stratum $\{S_i(v) = S_i(p) = 1\}$, suppose the $Y_i(v)$ are identically distributed as $F_{(v)}^{\text{alw.inf}}(\cdot)$, and the $Y_i(p)$ are identically distributed as $F_{(p)}^{\text{alw.inf}}(\cdot)$. Then, any functional that measures a contrast of the distributions

$$F_{(v)}^{alw.inf}(y) \equiv \Pr\{Y_i(v) \le y \mid S_i(v) = S_i(p) = 1\}$$
$$F_{(p)}^{alw.inf}(y) \equiv \Pr\{Y_i(p) \le y \mid S_i(v) = S_i(p) = 1\} \qquad (1)$$

is a causal estimand. Based on (1), a null hypothesis for no causal effect of vaccination on viral load in the always-infected principal stratum can be expressed as

$$H_0 : F_{(v)}^{\text{alw.inf}}(y) = F_{(p)}^{\text{alw.inf}}(y) \quad \text{for all} \quad y. \qquad (2)$$

Unfortunately, because neither distribution in (1) is identifiable (since $S_i(v)$ and $S_i(p)$ are not both observed), it is not possible to test (2) without introducing assumptions. In addition to SUTVA, two assumptions are useful for identifying the distributions:

A1: The assignment Z_i of each subject is independent of his/her potential outcomes.

A2: For each subject i, $\Pr\{S_i(v) = 1, S_i(p) = 0\} = 0$.

Assumption A1 plausibly holds in HIV vaccine efficacy trials due to randomization and blinding. A2 states that no subject would be infected if randomized to vaccine, but would be uninfected if randomized to placebo and, under A1, A2 will hold if vaccination does not increase the per-exposure infection probability for any subject. The SUTVA assumption may not hold, because HIV disease is infectious (Halloran and Struchiner, 1995); however if the study population is a small sample from a large population of susceptible individuals and there are few infectious contacts between trial participants, then it should approximately hold. SUTVA can be checked through epidemiologic studies and data on mixing of risk behavior among trial participants. Given SUTVA, A1 can be tested based on risk behavior data; under SUTVA and A1, A2 can be checked by testing if the HIV infection rate is higher in vaccine than placebo recipients for any participant subgroup.

Assumption A2 is very useful, because it implies that infected vaccine recipients must be in the always-infected principal stratum. Together with A1, this result implies that $F_{(v)}^{\text{alw.inf}}(y) = F_v(y) \equiv \Pr(Y_i^{\text{obs}} \le y \mid S_i^{\text{obs}} = 1, Z_i = v)$, where $F_v(\cdot)$ is the distribution of viral load in infected vaccine recipients; thus $F_{(v)}^{\text{alw.inf}}(\cdot)$ is identified from the observed data. A2 is similar to Angrist, Imbens, and Rubin's (1996) Monotonicity Assumption 5, which is useful for identifying a causal estimand defined for a principal stratum of compliers. On the other hand, A1 and A2 do not identify $F_{(p)}^{\text{alw.inf}}(\cdot)$, because they do not determine whether an infected placebo recipient is in the "protected" $\{S_i(v) = 0, S_i(p) = 1\}$ or always-infected $\{S_i(v) = 1, S_i(p) = 1\}$ stratum.

Given the randomization assignment and observed infection status of a trial participant, Table 1 indicates the principal stratum or strata to which the participant must belong, and lists the information available on potential viral loads. The table makes clear that the always-infected stratum is the natural subpopulation for causal inference on viral load, because it is the only stratum for which causal estimands involve only well-defined potential viral loads. Rubin (2000) made this point through a parallel example in which there are two randomized treatments, and vital status is observed one year after randomization, with the goal of causal inference being to assess the treatment effect on quality of life within the principal stratum of subjects alive under either treatment assignment.

2.2. *Logistic Selection Bias Models that Identify the Causal Estimands*

The set of subjects infected under randomization to placebo, $\{S_i(p) = 1\}$, partitions into the principal strata of protected and always-infected subjects, with the level of vaccine efficacy (VE) against infection determining the proportion in each. Specifically, define $\text{VE} = 1 - \text{RR} = 1 - \Pr\{S_i(v) = 1\}/\Pr\{S_i(p) = 1\}$; VE is a causal estimand measuring the relative reduction in infection risk conferred by randomizing to vaccine versus placebo. A2 implies $\text{VE} = \Pr\{S_i(v) = 0 \mid S_i(p) = 1\}$, which is the probability that a subject in $\{S_i(p) = 1\}$ is in the protected principal stratum (note that A2 is crucial here; $\Pr\{S_i(v) = 0 \mid S_i(p) = 1\}$ is not identified by randomization alone). The density of $Y(p)$ in subjects infected under randomization to placebo ($f_{(p)}(y)$) can be written as a mixture of the densities of $Y(p)$ for the protected ($f_{(p)}^{\text{prot}}(y)$) and always-infected ($f_{(p)}^{\text{alw.inf}}(y)$) strata:

$$f_{(p)}(y) = VE * f_{(p)}^{\text{prot}}(y) + (1 - \text{VE}) * f_{(p)}^{\text{alw.inf}}(y). \qquad (3)$$

Table 1
For the two randomization assignments $Z_i = v$, p, and infection outcomes $S_i^{obs} \equiv S_i(Z_i) = 0, 1$, the table indicates the basic principal stratum or strata to which the subjects belong, and the information available on the potential viral loads $Y_i(v)$ and $Y_i(p)$. Note that $Y_i(Z)$ is defined if and only if $S_i(Z) = 1$, $Z = v$, p, and the principal strata of uninfected placebo recipients and of infected vaccine recipients are known by assumption A2.

Randomized assign. Z_i	Observed infection status S_i^{obs}	Principal Stratum $\{S_i(v), S_i(p)\}$ and information on potential viral loads $Y_i(v)$, $Y_i(p)$		
Vaccine	Uninfected	**Protected** $\{S_i(v)=0,\ S_i(p)=1\}$ $Y_i(v)$ undefined $Y_i(p)$ unobserved	or	**Never-infected** $\{S_i(v)=0,\ S_i(p)=0\}$ $Y_i(v)$ undefined $Y_i(p)$ undefined
Placebo	Uninfected	$\{S_i(v)=1,\ S_i(p)=0\}$ (empty set by A2)		**Never-infected** $\{S_i(v)=0,\ S_i(p)=0\}$ $Y_i(v)$ undefined $Y_i(p)$ undefined
Vaccine	Infected	$\{S_i(v)=1,\ S_i(p)=0\}$ (empty set by A2)		**Always-infected** $\{S_i(v)=1,\ S_i(p)=1\}$ $Y_i(v)$ observed $Y_i(p)$ unobserved
Placebo	Infected	**Protected** $\{S_i(v)=0,\ S_i(p)=1\}$ $Y_i(v)$ undefined $Y_i(p)$ observed	or	**Always-infected** $\{S_i(v)=1,\ S_i(p)=1\}$ $Y_i(v)$ unobserved $Y_i(p)$ observed

With some calculations, the mixture (3) can be reexpressed as a biased sampling model

$$f_{(p)}^{\text{alw.inf}}(y) = W^{-1} w(y) f_{(p)}(y), \tag{4}$$

where $w(y) = \Pr\{S_i(v) = 1 \mid Y_i(p) = y,\ S_i(p) = 1\}$ and $W = \int_{-\infty}^{\infty} w(y) f_{(p)}(y)\, dy$ is a normalizing constant equal to $1 - \text{VE} = \text{RR}$. The weight function $w(y) = \text{RR}(y) = 1 - \text{VE}(y)$ is the probability that a subject infected with viral load y if randomized to placebo would be infected if randomized to vaccine.

Let $F_p(y)$ and $f_p(y)$ be the distribution and density of the viral load in infected placebo recipients. Under A1, $F_{(p)}(y) = F_p(y)$, and (4) can be restated as $f_{(p)}^{\text{alw.inf}}(y) = (1 - \text{VE})^{-1} w(y) f_p(y)$. Then, under A1–A2, the null hypothesis of interest (2) is equivalent to

$$H_0 : F_v(y) = (1 - \text{VE})^{-1} \int_{-\infty}^{y} w(z)\ dF_p(z) \quad \text{for all} \quad y. \tag{5}$$

By A1, VE is identified from the observed data. If $w(\cdot)$ were known, then both $F_{(v)}^{\text{alw.inf}}(\cdot)$ and $F_{(p)}^{\text{alw.inf}}(\cdot)$ would be identified, and the hypothesis (2) could be tested. However, $w(\cdot)$ is unknown, and it is not possible to test whether a particular $w(\cdot)$ is correctly specified from the data plus A1-A2. Our approach to this problem assumes $w(\cdot)$ is known, and tests (5) for a variety of fixed choices of $w(\cdot)$. For such an approach to be fruitful, it is important that the unidentified sensitivity function $w(\cdot)$ be interpretable.

Towards this goal, we parameterize $w(\cdot)$ as logistic, indexed by an interpretable selection bias parameter β, which allows $w(\cdot)$ to be constant or smoothly monotone increasing or decreasing: $w(y) = w(y \mid \alpha, \beta) = \exp(\alpha + \beta y) / \{1 + \exp(\alpha + \beta y)\}$. For β finite, e^{β} is the odds ratio of infection under randomization to vaccine given infection under randomization to placebo with viral load y versus with viral load $y - 1$. This interpretation allows the choice of β to be guided by beliefs about plausible degrees of selection bias. For fixed $\beta \in [-\infty, \infty]$, the logistic selection bias model is specified by

$$\begin{aligned} F_{(p)}^{\text{alw.inf}}(y) &= (1 - \text{VE})^{-1} \int_{-\infty}^{y} \frac{\exp(\alpha + \beta z)}{1 + \exp(\alpha + \beta z)}\ dF_p(z) \\ &\equiv F_p(y \mid \beta). \end{aligned} \tag{6}$$

Given fixed β, α is determined as the solution to the equation $F_p(\infty \mid \beta) = 1$.

Figure 1 illustrates five selection models specified by (6) and β fixed at $-\infty$, -1, 0, 1, or ∞, which represent different ways to distribute VE of the mass of $f_{(p)}(y)$ into the protected principal stratum via (3). Note that if $\text{VE} = 0$, there is no selection bias, regardless of β, and the higher VE, the greater the opportunity for bias. Heuristically, β specifies how much bias occurs through VE. Fixing $\beta = 0$ specifies a constant weight $w(y \mid \alpha,\ \beta = 0) = \text{RR}$, and reflects an assumption of no selection bias. Fixing $\beta > 0$ makes $w(y \mid \alpha,\ \beta)$ monotone increasing in y and reflects "positive" selection bias, with infection odds under randomization to vaccine higher for a larger PVL $Y(p) = y$. In this case, if the causal null hypothesis (2) is true, then the net vaccine effect is that $F_v(\cdot)$ is stochastically larger than $F_p(\cdot)$. Similarly, $\beta < 0$ makes $w(y \mid \alpha, \beta)$ monotone decreasing in y and reflects "negative" selection bias, with infection odds under randomization to vaccine lower for a

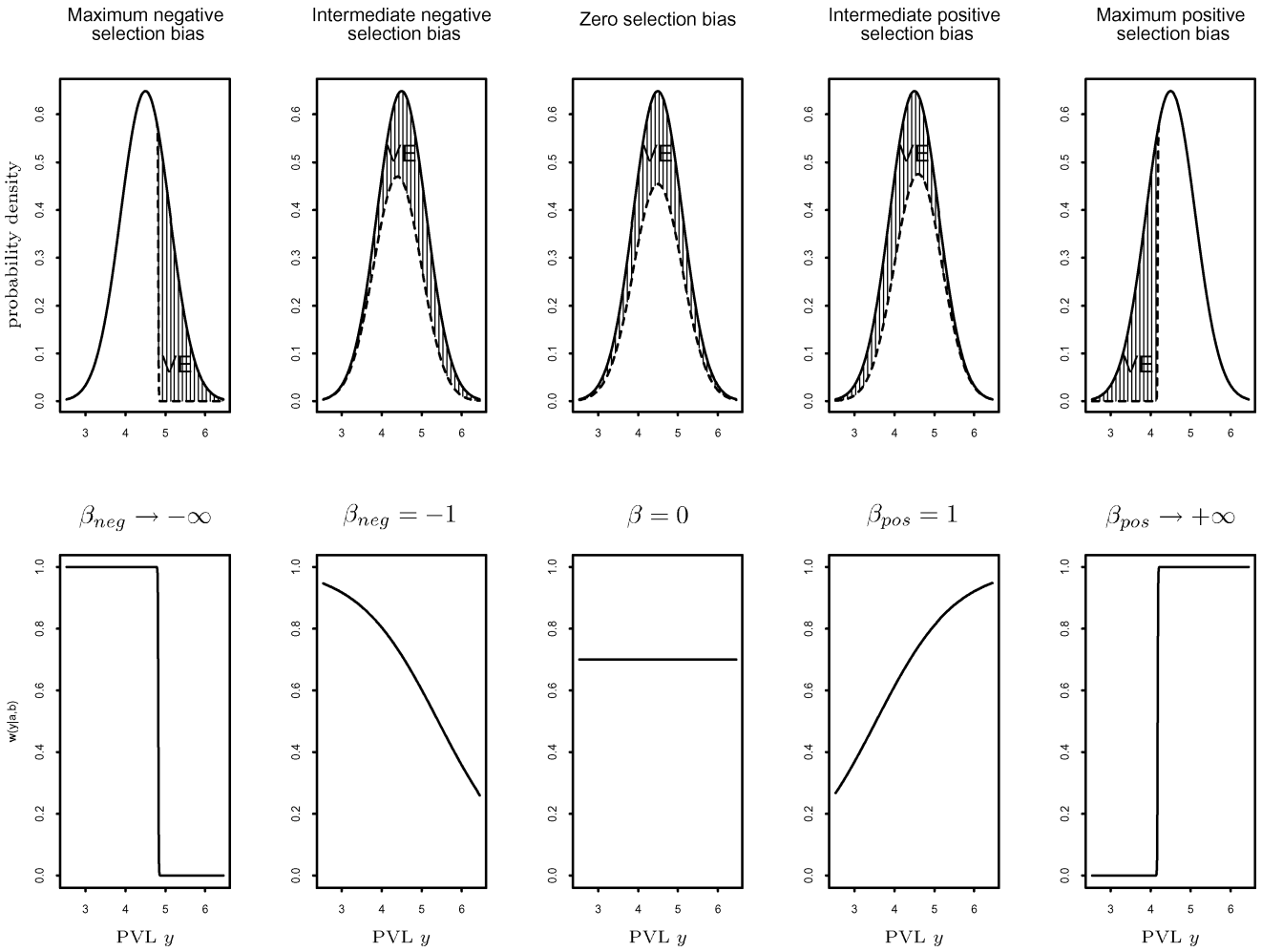

Figure 1. The upper panels show plots of the density of $Y_i(p)$ for subjects infected under randomization to placebo $\{S_i(p)=1\}$ (total area) partitioned into the subdensity for "protected" subjects not infected under randomization to vaccine $\{S_i(v)=0,\ S_i(p)=1\}$ (hatchmarked area $=$VE) and the subdensity for subjects "always infected" under randomization to either vaccine or placebo $\{S_i(v)=1,\ S_i(p)=1\}$ (unshaded area $=1-$VE). Using model (6) with $\beta=-\infty$, -1, 0, 1, or ∞, the five panels reflect different assumptions about how the vaccine relative risk $w(y\,|\,\alpha,\,\beta)=\mathrm{RR}(y)=\Pr\{S_i(v)=1\,|\,Y_i(p)=y,\ S_i(p)=1\}$ depends on the potential viral load (PVL) $Y_i(p)=y$ for subjects infected under randomization to placebo. The lower panels show corresponding plots of the logistic weight function $w(y\,|\,\alpha,\,\beta)$. The hatchmarked areas equal VE $=0.30$, and α was calculated from $1-\mathrm{VE}=\int_{-\infty}^{\infty} w(z\mid\alpha,\beta)dF_{(p)}(z)$ with $F_{(p)}(\cdot)$ given a normal distribution.

larger y; under (2) the net vaccine effect is that $F_v(\cdot)$ is stochastically smaller than $F_p(\cdot)$.

HHS developed tests for (5), using two models representing maximum plausible positive and negative bias. HHS's "positive" selection model is specified by placing all subjects in $\{S_i(p)=1\}$ with $Y(p)$ less than the VEth percentile $q^{\mathrm{VE}}_{(p)}$ of its distribution into the protected principal stratum; the "negative" selection model is specified by placing all subjects in $\{S_i(p)=1\}$ with $Y(p)$ greater than the upper VEth percentile $q^{1-\mathrm{VE}}_{(p)}$ of its distribution into the protected principal stratum. These models are limiting members of the class of logistic models (6), specified, respectively, by $\beta=\infty$ (Figure 1, rightmost panel) and $\beta=-\infty$ (Figure 1, left-most panel). To see this, note that setting $\alpha=-\beta q^{\mathrm{VE}}_{(p)}$ implies $\lim_{\beta\to\infty} w(y\,|\,\alpha,\,\beta)$ equals $I\{y>q^{\mathrm{VE}}_{(p)}\}$ for $y\neq q^{\mathrm{VE}}_{(p)}$ and $1/2$ for $y=q^{\mathrm{VE}}_{(p)}$, and that setting $\alpha=-\beta q^{1-\mathrm{VE}}_{(p)}$ implies $\lim_{\beta\to-\infty} w(y\,|\,\alpha,\,\beta)=I\{y<q^{1-\mathrm{VE}}_{(p)}\}$ for $y\neq q^{1-\mathrm{VE}}_{(p)}$ and $1/2$ for $y=q^{1-\mathrm{VE}}_{(p)}$. Therefore, based on the logistic weight function with β ranging between $-\infty$ to ∞, the class of models (6) spans all plausible magnitudes of selection bias.

If selection bias is presumed to follow model (6) for some unknown β within a plausible range β_{neg} to β_{pos}, then a two-sided null hypothesis representing no causal vaccine effect in the always-infected stratum, allowing for possible selection bias, is given by

$$H_{0\beta_{\mathrm{pos}},\beta_{\mathrm{neg}}}: F_p(\cdot\mid\beta_{\mathrm{pos}})\le F_v(\cdot)\le F_p(\cdot\mid\beta_{\mathrm{neg}}),\quad \beta_{\mathrm{pos}}\in[0,\infty],\beta_{\mathrm{neg}}\in[-\infty,0]. \tag{7}$$

Under A1–A2, $H_{0\beta_{\mathrm{pos}},\beta_{\mathrm{neg}}}$ is equivalent to $F^{\mathrm{alw.inf}}_{(p)}(\cdot)\le F^{\mathrm{alw.inf}}_{(v)}(\cdot)$, assuming model (6) with $\beta=\beta_{\mathrm{pos}}$ and $F^{\mathrm{alw.inf}}_{(v)}(\cdot)\le F^{\mathrm{alw.inf}}_{(p)}(\cdot)$, assuming model (6) with $\beta=\beta_{\mathrm{neg}}$. For the special

case $\beta_{\text{pos}} = \beta_{\text{neg}} = 0$, (7) collapses to the null hypothesis of no net vaccine effect on viral load, $H_0 : F_v(\cdot) = F_p(\cdot)$. Therefore, under the assumption of no selection bias, a standard comparison of viral load distributions between infected subgroups assesses the causal effect of vaccine in the always-infected principal stratum.

One-sided null hypotheses representing no causal vaccine effect are given by

$$H_{0\beta_{\text{pos}}} : F_p(\cdot \mid \beta_{\text{pos}}) \le F_v(\cdot), \qquad \beta_{\text{pos}} \in [0, \infty], \tag{8}$$

$$H_{0\beta_{\text{neg}}} : F_v(\cdot) \le F_p(\cdot \mid \beta_{\text{neg}}), \qquad \beta_{\text{neg}} \in [-\infty, 0]. \tag{9}$$

If (8) is rejected, then always-infected individuals have significantly higher viral loads under randomization to vaccine than placebo when controlling for selection bias. In sum, A1–A2 and model (6) can be used to specify a one- or two-sided null hypothesis for no causal effect of vaccine in the always-infected stratum that can be tested, and a sensitivity analysis can be performed by testing the hypothesis for a range of fixed values of β_{pos}, β_{neg}.

3. Statistical Hypothesis Tests and Estimation

Let $Y_{v1}, \ldots, Y_{vn_v}$ and $Y_{p1}, \ldots, Y_{pn_p}$ denote the viral loads of infected vaccine and placebo recipients, respectively. Y could be the average of two or more viral load measurements taken from an infected subject, or another continuous outcome such as the area under the longitudinal viral load curve. We assume each sample is independently, identically distributed, and the two samples are independent of one another.

3.1. *Nonparametric Test Statistics*

Fix $\beta_{\text{pos}} \ge 0$ and $\beta_{\text{neg}} \le 0$. Using the empirical distributions $\widehat{F}_v$ and $\widehat{F}_p$ calculated from the two observed samples, and an estimate of VE, nonparametric tests of $H_{0\beta_{\text{pos}}}$, $H_{0\beta_{\text{neg}}}$, and $H_{0\beta_{\text{pos}},\beta_{\text{neg}}}$ can be based on comparisons of $\widehat{F}_v(\cdot)$ with $\widehat{F}_p(\cdot \mid \beta_{\text{pos}})$, $\widehat{F}_p(\cdot \mid \beta_{\text{neg}})$, and both estimates, respectively. The VE parameter can be estimated by $\widehat{\text{VE}} = 1 - (n_v/N_v)/(n_p/N_p)$, with $N_v(N_p)$ the number of subjects randomized to vaccine (placebo). Under A1, $\widehat{\text{VE}}$ is unbiased for VE if the vaccine protects by an "all-or-none" mechanism, and is approximately unbiased if it protects by another mechanism, since HIV infection is a rare event (Halloran, Haber, and Longini, 1992). We consider three criterion functions for summarizing the comparisons, based on means, suprema, and integrated squared differences.

The statistic $T_{M\beta}$ for comparing means, appropriate for testing (8) or (9), is given by

$$T_{M\beta} = \int_{-\infty}^{\infty} y\{d\widehat{F}_v(y) - d\widehat{F}_p(y \mid \beta)\}, \tag{10}$$

where $\int_{-\infty}^{\infty} y\, d\widehat{F}_v(y) = n_v^{-1} \sum_{i=1}^{n_v} Y_{vi}$ and $\widehat{F}_p(y \mid \beta)$ is the maximum likelihood estimator of $F_p(y \mid \beta)$ under model (6), calculated as

$$\widehat{F}_p(y \mid \beta) = (1 - \widehat{\text{VE}})^{-1} \frac{1}{n_p} \sum_{i=1}^{n_p} I\{Y_{pi} \le y\}\, w(Y_{pi} \mid \widehat{\alpha}, \beta).$$

Here, $\widehat{\alpha}$ is computed by solving the equation $\widehat{F}_p(\infty \mid \beta) = 1$ for α, i.e., α solves

$$1 - \widehat{\text{VE}} = \int_{-\infty}^{\infty} \frac{\exp(\alpha + \beta y)}{1 + \exp(\alpha + \beta y)}\, d\widehat{F}_p(y). \tag{11}$$

A solution to (11) can be found rapidly using a numerical one-dimensional line search.

The null hypothesis $H_{0\beta_{\text{pos}}}$ is rejected if $T_{M\beta_{\text{pos}}}$ is large. For large positive β_{pos} and $\widehat{\alpha} = -\beta_{\text{pos}} \widehat{q}_p^{\widehat{\text{VE}}}$, with $\widehat{q}_p^{\widehat{\text{VE}}}$ the $\{n_p(1 - \widehat{\text{VE}})\}$th largest value of $Y_{p1}, \ldots, Y_{pn_p}$, $T_{M\beta_{\text{pos}}}$ reduces to HHS's nonparametric statistic T_M that tests (8) with $\beta_{\text{pos}} = \infty$. Similarly, $H_{0\beta_{\text{neg}}}$ is rejected if $T_{M\beta_{\text{neg}}}$ is negative and large, and for large negative β_{neg} and $\widehat{\alpha} = -\beta_{\text{neg}} \widehat{q}_p^{1-\widehat{\text{VE}}}$, $T_{M\beta_{\text{neg}}}$ reduces to HHS's statistic T_M that tests (9) with $\beta_{\text{neg}} = -\infty$. The maximum of $|T_{M\beta_{\text{pos}}}|$ and $|T_{M\beta_{\text{neg}}}|$ can be used for a two-sided test of (7).

Second, a one-sided Kolmogorov-Smirnov-type statistic for testing (8) is defined by

$$T_{\text{KS}\beta_{\text{pos}}} = m^{1/2} \text{sup}_{-\infty < y < \infty} |\{\widehat{F}_p(y \mid \beta_{\text{pos}}) - \widehat{F}_v(y)\} \vee 0|, \tag{12}$$

with $n = n_v + n_p$, $m = (n_v n_p)/n$; an Anderson-Darling-type statistic is defined by

$$T_{\text{AD}\beta_{\text{pos}}} = m \int_{-\infty}^{\infty} \frac{[\{\widehat{F}_p(y \mid \beta_{\text{pos}}) - \widehat{F}_v(y)\} \vee 0]^2}{\widehat{H}_n(y \mid \beta_{\text{pos}})(1 - \widehat{H}_n(y \mid \beta_{\text{pos}}))} \times d\widehat{H}_n(y \mid \beta_{\text{pos}}), \tag{13}$$

where $\widehat{H}_n(y \mid \beta_{\text{pos}}) = (n_p/n)\widehat{F}_p(y \mid \beta_{\text{pos}}) + (n_v/n)\widehat{F}_v(y)$. One-sided statistics for testing (9) are given by (12) and (13) with β_{pos} replaced by β_{neg} and $\vee$ replaced by $\wedge$. Two-sided statistics for testing (7) can be defined similarly. When $\beta = \beta_{\text{pos}} = \beta_{\text{neg}} = 0$, the two-sided statistics reduce to the classical Kolmogorov-Smirnov and Anderson-Darling test statistics for comparing two distribution functions (D'Agostino and Stephens, 1986).

3.2. *Computing Critical Values for the Tests*

We use a modification of the "controls only" bootstrap procedure developed by HHS for computing critical values for the test statistics. The modification is that once the bootstrap estimate of vaccine efficacy $\widehat{\text{VE}}^{\star}$ is computed as in HHS, a bootstrap estimate $\widehat{\alpha}^{\star}$ is computed as the solution to equation (11), with $\widehat{\text{VE}}$ replaced by $\widehat{\text{VE}}^{\star}$. Estimating VE within each bootstrap iteration appropriately accounts for the uncertainty in $\widehat{\text{VE}}$.

The nonparametric bootstrap tends to approximate smooth distributions better than distributions with discontinuities. For β finite, the use of a smooth logistic selection weight function in the test statistics suggests that the nonparametric bootstrap should perform well. For $|\beta|$ infinite, $F_p(\cdot \mid \beta_{\text{pos}})$ and $F_p(\cdot \mid \beta_{\text{neg}})$ have discontinuities at the truncation point, which could abrogate bootstrap performance. The simulation study confirms that tests with $\beta_{\text{pos}} = \infty$ have poorer size and power characteristics than tests with $\beta_{\text{pos}} = 1$.

3.2. *Nonparametric Estimation*

Under A1–A2 and a model (6) with β fixed, $\widehat{\mu}_{\text{ACE}}(\beta) \equiv -T_{M\beta}$ is a consistent estimate of the *average causal effect* (ACE) parameter $\mu_{\text{ACE}}(\beta) = \int_{-\infty}^{\infty} y\{dF_{(p)}^{\text{alw.inf}}(y) - dF_{(v)}^{\text{alw.inf}}(y)\}$. By

bootstrap resampling from $\widehat{F}_v(y)$ and $\widehat{F}_p(y \mid \beta)$, 95% bootstrap percentile confidence intervals about $\mu_{\text{ACE}}(\beta)$ can be constructed. An estimation-based sensitivity analysis can be carried out by plotting point and interval estimates of $\mu_{\text{ACE}}(\beta)$ versus β.

4. Simulation Study

Through simulations of an HIV vaccine trial, we evaluate the three one-sided tests of the null hypothesis $H_{0\beta_{\text{pos}}}$ in (8). Rejecting (8) implies that individuals infected under either assignment have significantly higher viral load if assigned vaccine than if assigned placebo when controlling for selection bias specified by β_{pos} and model (6). We consider an intermediate-sized efficacy trial with 45 infections expected in the placebo group (Rida et al., 1997), and suppose the true VE equals 30% or 50%. The true amount of selection bias is determined by the parameter $\beta = \beta_{\text{pos}}$ in model (6), with $\beta = 0$, 1, or ∞. Thus, data are generated under three kinds of null models, which assume no selection bias, an intermediate amount of selection bias (supposing the infection odds under vaccine of a subject who would be infected under placebo increases $e^1 = 2.72$-fold per one unit higher PVL $Y(p)$), and maximal plausible positive selection bias. We assume two independent measures of viral load are available per person. The sample $Y_{p1}, \ldots, Y_{pn_p}$ is generated from a normal distribution $F_p(y)$ with mean 4.50 and variance 0.36. These parameter values equal those used by HHS, selected based on a cohort of recently HIV-infected people. The sample $Y_{v1}, \ldots, Y_{vn_v}$ is generated from $F_p(\cdot \mid \beta)$ in model (6) with true $\beta = 0$, 1, or ∞. For each true β, three vaccine effects on viral load in the always-infected are evaluated: mean shifts of $\Delta = 0, 1/3,$ or $1/2$ $\log_{10}$; these are over and above any selection bias induced by the true β, i.e., the samples are drawn such that $F_{(p)}^{\text{alw.inf}}(y) = F_p(y \mid \beta)$ and $F_{(v)}^{\text{alw.inf}}(y) = F_p(y - \Delta \mid \beta)$.

For each of 500 datasets simulated under each parameter configuration, the three one-sided test statistics are calculated, with presumed selection bias levels $\beta = 0$, 1, or ∞. Critical values for the tests are determined using 500 bootstrap replications.

Using a nominal 5% type I error level, Table 2 shows estimated sizes and powers of the tests. The sizes are judged by the bolded rows, for which the correct amount of selection bias is presumed (true β = presumed β). All tests have empirical size close to nominal, except when VE = 30% and $\beta = \infty$; then the size is inflated to 8–12%. The elevated size is caused by the simulated trials with estimated VE less than zero; this occurred 18 times and of these, the nonparametric mean-based test rejected the null hypothesis 16 times. If the 18 trials with $\widehat{\text{VE}} < 0$ are discarded, then the rejection rate is 5.8%. A similar pattern was seen for the other test statistics. When $\widehat{\text{VE}} < 0$, the testing procedure operates under the assumption of no selection bias, and simply tests H_0: $F_v(\cdot) = F_p(\cdot)$. However, in fact $F_v(\cdot) < F_p(\cdot)$ due to positive selection bias (true $\beta = \infty$), which explains the inflated probability of rejection. Note that when VE = 50%, the sizes are not elevated, because the estimated VE is rarely negative. In summary, the tests generally have nominal size, except that when VE is low and the sample size is moderate, underestimation of VE can lead to an increased risk of false rejection. This phenomenon was also found by HHS.

When the correct selection bias model is assumed, the three tests have comparable power, with that of the Kolmogorov-Smirnov procedure being slightly less. In addition, power diminishes as the true β increases. Next we consider power when

Table 2

Power × 100% *for detecting a* 0, 1/3, *and* 1/2 $\log_{10}$ *mean-shift alternative, over and above any selection bias induced by the true* β*, based on a one-sided* 5% *level test*

True	Presumed	Nonparametric mean			Kolmogorov-Smirnov			Anderson-Darling		
β	β	0	1/3	1/2	0	1/3	1/2	0	1/3	1/2
					VE = 30%					
0	0	**4.8**	**73.8**	**96.8**	**5.4**	**69.0**	**91.2**	**4.0**	**72.0**	**94.4**
0	1	3.2	49.8	82.0	3.0	44.8	78.2	2.4	46.8	79.6
0	∞	0.2	15.8	36.0	0.8	16.8	37.6	0.2	15.4	33.6
1	0	14.6	91.8	99.8	15.6	85.4	97.8	14.4	90.2	99.2
1	1	**5.8**	**67.6**	**94.6**	**7.2**	**63.4**	**90.8**	**5.4**	**65.4**	**94.0**
1	∞	1.8	26.0	52.2	1.4	27.6	51.8	1.6	25.8	48.8
∞	0	69.6	100	100	76.2	100	100	79.8	100	100
∞	1	37.4	95.8	100	42.6	97.0	100	45.8	97.4	100
∞	∞	**8.8**	**55.0**	**83.2**	**12.0**	**55.2**	**82.4**	**10.8**	**57.2**	**86.2**
					VE = 50%					
0	0	**6.6**	**70.0**	**93.2**	**5.8**	**63.8**	**89.8**	**5.8**	**66.4**	**92.6**
0	1	1.0	30.6	64.2	0.4	25.8	58.0	0.2	26.6	61.2
0	∞	0.2	3.4	10.2	0.4	3.6	9.0	0.2	2.8	7.4
1	0	25.8	92.2	99.0	24.8	87.4	99.0	24.0	91.0	99.2
1	1	**5.0**	**62.4**	**88.2**	**5.8**	**57.2**	**84.2**	**4.2**	**60.6**	**87.8**
1	∞	0.4	9.8	30.6	0.2	8.6	25.0	0.2	8.2	24.6
∞	0	94.8	100	100	97.2	100	100	97.6	100	100
∞	1	56.8	99.2	100	72.6	99.8	100	69.0	99.6	100
∞	∞	**5.2**	**59.2**	**84.2**	**6.0**	**55.6**	**79.4**	**6.4**	**57.8**	**82.0**

an incorrect amount of selection bias is presumed (Table 2, unbolded rows). If zero bias is presumed ($\beta = 0$), but in truth there is moderate bias ($\beta = 1$), power is high, but at the expense of an inflated false rejection rate; it is at 15% when VE = 30% and 26% when VE = 50%. If zero bias is presumed and there is actually extreme bias ($\beta = \infty$), then power is extremely high and the sizes are extremely inflated. This illustrates the importance of accounting for the possibility of selection bias to avoid being misled. Next, suppose there is no selection bias, but one conservatively presumes $\beta = \infty$. Then power drops severely, e.g., to 10% for detecting a 1/2 $\log_{10}$ mean shift when VE = 50%, compared to 93% if the correct $\beta = 0$ is assumed. For the more moderate assumption $\beta = 1$, a much smaller price is paid, with power dropping to 64%. Thus, making a highly conservative assumption of maximal selection bias can cause great power loss. This finding supports the use of a continuously indexed sensitivity analysis, as proposed here.

The Kolmogorov-Smirnov-type and Anderson-Darling-type tests are expected to have greater power than the mean-based test for detecting non-mean-shift alternatives. We studied this conjecture by generating placebo group viral loads from a normal mixture distribution $0.5N(3.50, 0.36) + 0.5\ N(5.50, 0.36)$ (e.g., infection with a mild or virulent virus) and vaccine group viral loads from a mixture of truncated normal distributions $0.5\,\text{TruncN}(3.50, 0.36) + 0.5\,\text{TruncN}(5.50, 0.81)$, with truncation point of each distribution at the 70th percentile. Presuming $\beta = \infty$, the Kolmogorov-Smirnov-type test has 78% power to reject (8), while the other tests have between 13% and 20% power. Thus, if the viral load distributions are expected to differ in respects other than a mean-shift, then the Kolmogorov-Smirnov-type and Anderson-Darling-type tests merit consideration.

5. Example

To illustrate how a sensitivity analysis could be carried out on a forthcoming vaccine trial dataset, we analyze a single dataset, simulated using Gaussian distributions assuming $n_p = 45$ infections in the placebo group, $\widehat{\text{VE}} = 40\%$ (and thus $n_v = 27$ infections in the vaccine group), a true causal vaccine effect to reduce the mean viral load in the always-infected by $0.33\log_{10}$, and true $\beta = \beta_{\text{neg}} = -1$, i.e., moderate negative selection bias that leads to lower viral loads in infected vaccine recipients. The true causal and biasing vaccine effects on viral load imply that the net vaccine effect on mean viral load is $0.49\log_{10}$. For β_{neg} ranging in $[-\infty, 0]$, we consider testing $H_{0\beta_{\text{neg}}}$ in (9) versus the alternative hypothesis that vaccination lowers viral load in the always-infected.

The first step is to produce descriptive plots and summary measures comparing the observed viral load distributions between the infected subgroups. The average viral loads are 3.96 and 4.48 in the infected vaccine and placebo groups, respectively. The second step is to calculate a test statistic for values of $\beta = \beta_{\text{neg}}$ ranging between 0 and a negative value that makes the selection bias odds ratio $\text{OR} = e^{-\beta}$ large (e.g., $\beta = -5$ yields $e^{-\beta} = 148$), and for the extreme model ($\beta = -\infty$). The third step is to plot the p-value of the test statistic versus OR, which will always be monotone except for stochastic variations in the bootstrap. This provides a graphical sensitivity analysis (Figure 2). Fourth, calculation of the value of β at which the test statistic is exactly statistically significant at the 0.025 level allows one to assess the extent of selection bias needed to lose the significance of the result. A 0.025 significance level is chosen because the test is one-sided. In this example, the critical β value for the test statistic $T_{M\beta}$ is -1.83, which implies the selection odds ratio must be at least $e^{1.83} = 6.23$ before the significance of the result is lost. The sensitivity analyses based on the other two test statistics give similar results (Figure 2).

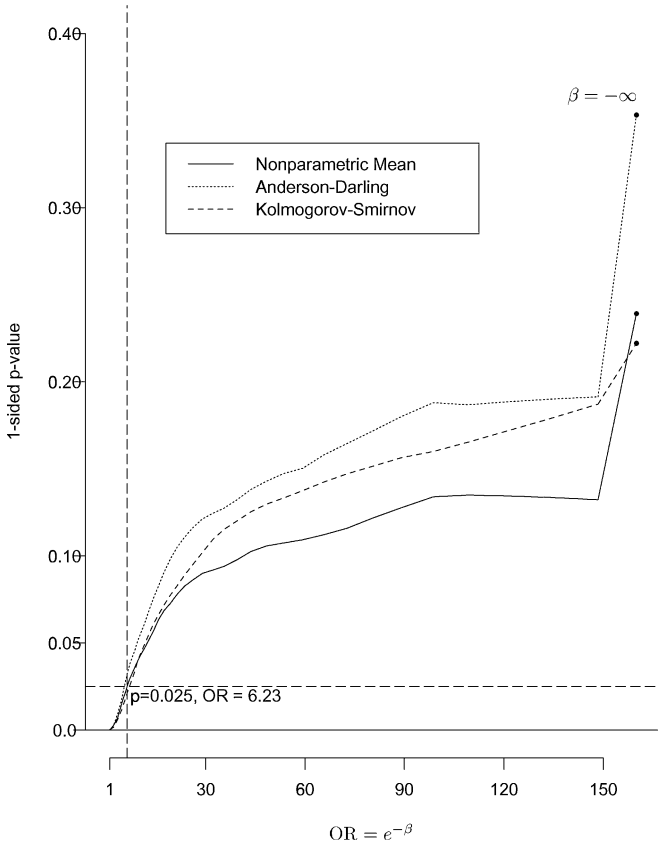

Figure 2. Based on the nonparametric mean-based, Anderson-Darling-type and Kolmogorov-Smirnov-type test statistics, the figure shows the one-sided bootstrap p-value plotted as a function of the selection bias odds ratio $\text{OR} = e^{-\beta} = e^{-\beta_{\text{neg}}}$; e^{β} is the odds ratio of infection under randomization to vaccine given infection under randomization to placebo with viral load y versus with viral load $y - 1$. If the magnitude of selection bias is believed to be less than $\text{OR} = e^{1.83} = 6.23$, then a significant causal effect of vaccination to lower viral load in the always-infected subpopulation can be inferred.

Fifth, an estimation-based sensitivity analysis can be carried out (Figure 3). Suppose vaccination must lower mean viral load in the always-infected by at least $0.2\log_{10}$ to be considered clinically significant. The value of β at which the lower 95% confidence limit for $\mu_{\text{ACE}}(\beta)$ crosses 0.2 is -0.50, corresponding to a "critical" odds ratio of $e^{0.50} = 1.65$.

Sixth, interpretations are made. In this example, a study team might conclude that it is unlikely that selection bias could fully explain the observed lower viral loads in infected vaccine recipients, and therefore a genuine viral suppressing

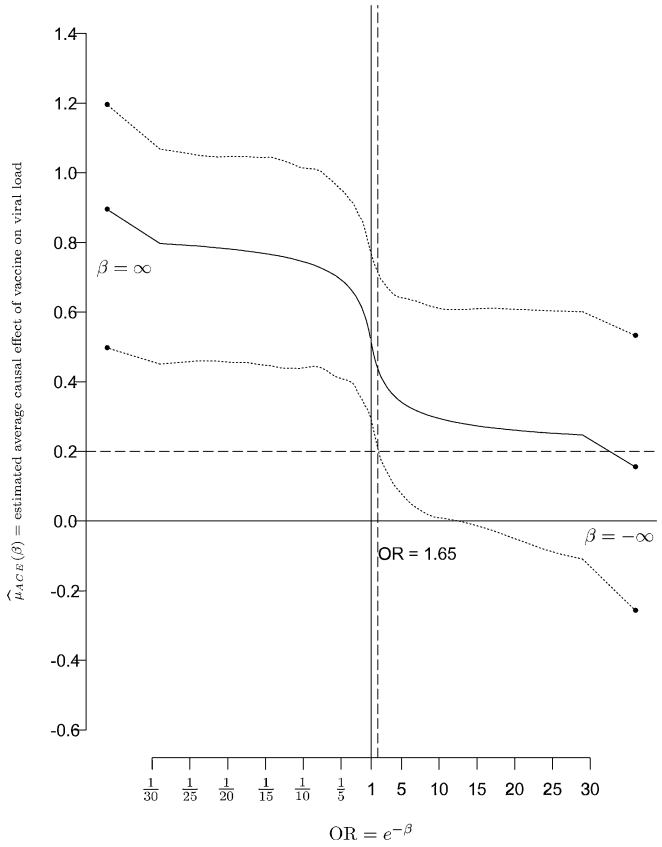

Figure 3. Point estimates $\widehat{\mu}_{\text{ACE}}(\beta) = -T_{M\beta}$ (bold line) and bootstrap 95% confidence intervals (dotted lines) for the average causal effect of vaccine $\mu_{\text{ACE}}(\beta) = \int_{-\infty}^{\infty} y\{dF_{(p)}^{\text{alw.inf}}(y) - dF_{(v)}^{\text{alw.inf}}(y)\}$ in the always-infected principal stratum as a function of the selection bias odds ratio $\text{OR} = e^{-\beta_{\text{pos}}}$ (left side of 0) and of $\text{OR} = e^{-\beta_{\text{neg}}}$ (right side of 0); e^{β} is the odds ratio of infection under randomization to vaccine given infection under randomization to placebo with viral load y versus with viral load $y - 1$. If the magnitude of selection bias is believed to be less than $\text{OR} = e^{0.50} = 1.65$, then a significant causal effect of vaccination to lower the mean viral load by at least $0.2 \log_{10}$ in the always-infected subpopulation can be inferred.

effect of vaccine in the always-infected is inferred. However, whether the effect is clinically significant is inconclusive. These conclusions would be based on beliefs that a selection bias effect with odds ratio 6.23 or higher is implausible, but a selection odds ratio of 1.65 is not unexpected.

Note that if only the hypothesis $H_{0\beta_{\text{neg}}}$ with extreme selection bias $\beta_{\text{neg}} = -\infty$ had been tested, then the team would likely not be able to conclude that vaccination reduced viral load in the always-infected (p-value > 0.20, Figure 2). This illustrates the added value of a continuously indexed sensitivity analysis.

6. Discussion

Appropriate interpretation of analyses of vaccine effects on viral load is challenging. Two main reasons are the lack of validation of viral load measures as accurate surrogates for secondary transmission and disease progression, and the potential for selective effects of the vaccine to bias inferences. As with HHS, we address the second problem, and extend their work to provide a method of sensitivity analysis over a continuous range of levels of putative selective effects. Since the true amount of selection bias may be considerably less than the worst-case amounts considered by HHS, the methods developed here may provide for more powerful assessments.

As illustrated in the Section 5, lower viral loads observed in infected vaccine recipients compared to infected placebo recipients could be caused partly by a causal viral suppressing effect of vaccine in the always-infected principal stratum and partly by selective vaccine protection against viruses that produce higher viral loads. Both effects are beneficial, and the assessment of the net vaccine effect (in the infected subgroups) gives useful information about the overall benefit of vaccination; this result should be reported together with the causal inference for the always-infected principal stratum. On the other hand, as for assessing a vaccine's effect of increasing viral load, the inference on the net vaccine effect could dangerously mislead. Selection bias could create higher viral loads in infected vaccine recipients compared to infected placebo recipients, i.e., produce a negative net vaccine effect, even though the vaccine has no adverse causal effect on viral load and has beneficial $\text{VE} > 0$. Therefore, it is crucial to build robustness to selection bias into assessments of vaccine harm, to protect against a spurious conclusion that could prevent use of or slow development of a safe and partially efficacious vaccine.

Within the framework of FR, this article develops techniques for causal inference in the always-infected principal stratum. Alternatively, causal inference could be made using a missing data framework that assumes all randomized subjects will eventually become HIV infected, and thus at some point will have a viral load value. In such an approach, the viral load is missing in subjects who have not yet been infected by the time of the analysis, and causal estimands can be defined based on functionals of contrasts of the viral load distributions for the vaccine and placebo groups. The goal of assessing such estimands is to compare the viral load distribution between the randomized groups had (contrary to fact) all subjects been infected during the trial. Rotnitzky and Robins (1997) developed an inverse probability of censoring weighted estimating equations method that could be used for causal inference on a mean-difference version of this estimand. This technique would model the viral load by a semiparametric conditional mean model with unspecified error distribution, and the infection probability (i.e., the response probability) by a parametric model. If the hazard rate of infection rather than the binary infection probability was modeled, then Scharfstein et al.'s (1999) method would apply for making inference on the same estimand. The advantages of these approaches include that they minimize modeling assumptions, that they can incorporate predictors of the infection risk, and that they can be used for sensitivity analysis of the effect of misspecification of the model for infection risk. The drawback of any such missing-data approach for the present application is that the causal estimand may not be relevant or interpretable, because it is unrealistic to suppose that all subjects would

eventually be HIV infected. FR criticize use of such a causal estimand because it uses nonexistent "*a priori*" counterfactuals. Inferences for the always-infected subpopulation provide interpretable and practical information for vaccine recipients who become HIV infected despite vaccination.

In addition to HIV vaccine trials, the methods developed here apply to general randomized clinical trials, for sensitivity analyses of causal treatment effects in the subpopulation of subjects who would experience a postrandomization event under either assignment.

Acknowledgements

The authors thank Andrea Rotnitzky, Steve Self, the associate editor, and referees for helpful comments. This work was supported by NIH grants AI46703-01 and AI38855.

Résumé

Des vaccins avec une capacité limitée d'empêcher l'infection HIV peuvent avoir un impact positif sur la pandémie HIV/SIDA en limitant une transmission secondaire et une maladie chez les receveurs qui deviennent infectés. Pour évaluer l'impact de la vaccination sur une transmission secondaire et sur la maladie, des essais d'efficacité estiment les effets du vaccin sur la charge virale du HIV et d'autres aspects ultérieurs mesurés après l'infection. Un test standard qui compare la distribution de la charge virale entre les sous-groupes infectés de vaccinés et les receveurs placebo n'estime pas un effet causal de la vaccination, car les groupes de comparaison sont sélectionnés après randomisation. Pour résoudre ce problème, nous formulons une estimation cliniquement adaptée de la causalité par l'utilisation du cadre de la stratification principale développée par Frangakis et Rubin (2002), et nous proposons une classe de modèles logistiques biaisés de sélection dont les membres identifient les estimations. Etant donné un modèle de sélection dans la classe, des procédures sont développées pour tester et estimer les effets causaux de la vaccination sur la charge virale dans la strate principale des sujets qui seraient infectés sans se soucier de l'affectation de randomisation. Nous montrons comment les procédures peuvent être utilisées pour une analyse de sensibilité qui quantifie comment l'effet causal de la vaccination varie avec la magnitude supposée du biais de sélection.

References

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

Barouch, D. H., Kunstman, J., Kuroda, M. J., et al. (2002). Eventual AIDS vaccine failure in a rhesus monkey by viral escape from cytotoxic T lymphocytes. *Nature* **415,** 335–339.

Burke, D. S. (1992). Human HIV vaccine trials: Does antibody-dependent enhancement pose a genuine risk? *Perspectives in Biology and Medicine* **35,** 511–530.

Coombs, R. W., Speck, C. E., Hughes, J. P., et al. (1998). Association between culturable human immunodeficiency virus type 1 (HIV-1) in semen and HIV-1 RNA levels in semen and blood: Evidence for compartmentalization of HIV-1 between semen and blood. *Journal of Infectious Diseases* **177,** 320–330.

D'Agostino, R. B. and Stephens, M. A. (1986). Goodness-of-fit techniques. New York: Marcel Dekker.

Francis, D. P., Gregory, T., McElrath, M. J., Belshe, R. B., Gorse, G. J., Migasena, S., Kitayaporn, D., Pitisuttitham, P., Matthews, T., Schwartz, D. H., and Berman, P. W. (1998). Advancing AIDSVAX to Phase III: Safety, immunogenicity, and plans for Phase III. *AIDS Research and Human Retroviruses,* **14**(Suppl. 3), S325–S331.

Frangakis, C. E. and Rubin, D. B. (2002). Principal stratification in causal inference. *Biometrics* **58,** 21–29.

Goetghebeur, E., Kenward, M., Molenberghs, G., and Vansteelandt, S. (2000). Inferential tools for sensitivity analysis and noncompliance in clinical trials. In *Proceedings of the Annual Meeting.* Indianapolis, Indiana: American Statistical Association.

Halloran, M. E. and Struchiner, C. J. (1995). Causal inference in infectious diseases. *Epidemiology* **6,** 142–151.

Halloran, M. E., Haber, M. J., and Longini, I. M. (1992). Interpretation and estimation of vaccine efficacy under heterogeneity. *American Journal of Epidemiology* **136,** 328–343.

Holland, P. (1986). Statistics and causal inference. *Journal of the American Statistical Association* **81,** 945–961.

Hudgens, M. G., Hoering, A., and Self, S. G. (2003). On the analysis of viral load endpoints in HIV vaccine trials. *Statistics in Medicine,* in press.

Mascola, J. R., Mathieson, B. J., Zack, P. M., Walker, M. C., Halstead, S. B., and Burke, D. S. (1993). Summary report: Workshop on the potential risks of antibody-dependent enhancement in human HIV vaccine trials. *AIDS Research and Human Retroviruses* **9,** 1175–1184.

Mellors, J. W., Munoz, A., Giorgi, J. V., et al. (1997). Plasma viral load and CD4+ lymphocytes as prognostic markers of HIV-1 infection. *Annals of Internal Medicine* **126,** 946–954.

Nabel, G. J. (2001). Challenges and opportunities for development of an AIDS vaccine. *Nature* **410,** 1002–1007.

Quinn, T. C., Wawer, M. J., Sewankambo, N., Serwadda, D., Li, C., Wabwire-Mangen, F., Meehan, M. O., Lutalo, T., and Gray, R. H. (2000). Viral load and heterosexual transmission of human immunodeficiency virus type 1. *New England Journal of Medicine* **342,** 921–929.

Rida, W. N., Fast, P., Hoff, R., and Fleming, T. R. (1997). Intermediate-sized trials for the evaluation of HIV vaccine candidates: A workshop summary. *Journal of Acquired Immune Deficiency Syndrome* **16,** 195–203.

Rida, W. N. and Lawrence, D. L. (1995). Prophylactic HIV vaccine trials. In *AIDS Clinical Trials: Guidelines for Design and Analysis,* D. M. Finkelstein, D. A. Schoenfeld (eds), 319–348. New York: Wiley-Liss.

Robins, J. M. and Greenland, S. (1992). Identifiability and exchangeability of direct and indirect effects. *Epidemiology* **3,** 143–155.

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.

Rosenbaum, P. R. and Rubin, D. B. (1983). Assessing sensitivity to an unobserved binary covariate in an

observational study with binary outcome. *Journal of the Royal Statistical Society, Series B* **45,** 212–218.

Rotnitzky, A. and Robins, J. (1997). Analysis of semiparametric regression models with nonignorable nonresponse. *Biometrics* **16,** 81–102.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* **66,** 688–701.

Rubin, D. B. (1978). Bayesian inference for causal effects. *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., and Robins, J. M. (1999). Adjusting for nonignorable drop-out using semiparametric nonresponse models. *Journal of the American Statistical Association* **94,** 1096–1146.

Shiver, J. W., Fu, T.-M., Chen, L., et al. (2002). Replication-incompetent adenoviral vaccine vector elicits effective anti-immunodeficiency virus immunity. *Nature* **415,** 331–335.

Sterling, T. R., Vlahov, D., Astemborski, J., Hoover, D. R., Margolick, J. B., and Quinn, T. C. (2001). Initial plasma HIV-1 RNA levels and progression to AIDS in women and men. *New England Journal of Medicine* **344,** 720–725.

Received April 2002. *Revised February* 2003.
Accepted February 2003.