Published in final edited form as:

Biometrics. 2011 September ; 67(3): 1100–1110. doi:10.1111/j.1541-0420.2010.01508.x.

Sensitivity Analyses Comparing Time-to-Event Outcomes Only Existing in a Subset Selected Postrandomization and Relaxing Monotonicity

Bryan E. Shepherd^{1,*}, Peter B. Gilbert², and Charles T. Dupont¹

¹Department of Biostatistics, Vanderbilt University School of Medicine, 1161 21st Avenue South, Nashville, TN 37232, USA

²Department of Biostatistics, University of Washington, Seattle, WA 98195, USA

Summary

In randomized studies researchers may be interested in the effect of treatment assignment on a time-to-event outcome that only exists in a subset selected after randomization. For example, in preventative HIV vaccine trials, it is of interest to determine whether randomization to vaccine affects the time from infection diagnosis until initiation of antiretroviral therapy. Earlier work assessed the effect of treatment on outcome among the principal stratum of individuals who would have been selected regardless of treatment assignment. These studies assumed monotonicity, that one of the principal strata was empty (eg, every person infected in the vaccine arm would have been infected if randomized to placebo). Here we present a sensitivity analysis approach for relaxing monotonicity with a time-to-event outcome. We also consider scenarios where selection is unknown for some subjects because of non-informative censoring (e.g., infection status k years after randomization is unknown for some because of staggered study entry). We illustrate our method using data from an HIV vaccine trial.

Keywords

Causal inference; HIV; Kaplan-Meier; Monotonicity; Principal stratification

1. Introduction

In randomized controlled trials, researchers are often interested in evaluating the effect of treatment in a subgroup of individuals selected post-randomization. For example, in HIV vaccine trials of T cell based vaccines, it is of primary interest to assess the effect of vaccination on outcomes only measured in those who get infected (eg, time from infection diagnosis to initiation of antiretroviral therapy), as candidate T cell based vaccines are specifically designed to control rather than prevent infection (Mehrotra, Li, and Gilbert 2006; Graham 2002; Buchbinder et al. 2008; www.hvtm.org).

*bryan.shepherd@vanderbilt.edu

Supplementary Materials: Figures referenced in Sections 6 and 7 are available under the Paper Information link at the Biometrics website <http://www.biometrics.tibs.org>. The animation referenced in Section 6 can be viewed with Firefox at <http://biostat.mc.vanderbilt.edu/AnimatedVaxGen>. Code used in this manuscript is available at <http://biostat.mc.vanderbilt.edu/CodeSensitivitySGD> and is implemented with the R function `sensitivitySGD` in the R package `sensitivityPStrat`.

When an analysis selects on a post-randomization variable such as infection, selection bias could arise. Even though treatment is randomly assigned, subjects selected based on a post-randomization variable are not randomly selected. Subject characteristics of those selected could differ between treatment arms if treatment assignment influences selection (Rosenbaum 1984).

In such situations, Frangakis and Rubin (2002) proposed focusing inference on the subgroup of individuals whose post-randomization selection variable is the same regardless of treatment assignment. These subgroups have been coined principal strata. For example, in the HIV vaccine studies, researchers have estimated the effect of vaccination on post-infection outcomes among those who would have been infected regardless of treatment assignment (Hudgens, Hoering, and Self 2003 (HHS); Gilbert, Bosch, and Hudgens 2003 (GBH)).

Unfortunately, principal stratum membership is not known because participants are only assigned to one treatment. Therefore, assumptions and/or sensitivity analyses are required to estimate causal effects within principal strata. HHS and GBH assumed monotonicity – that every person infected in the vaccine arm would have been infected had they been randomized to placebo. They then performed a sensitivity analysis examining results under different assumptions regarding the probability of infection if assigned vaccine given infection under placebo and the outcome. Shepherd, Gilbert, and Lumley (2007) (SGL) extended HHS and GBH to the situation where the outcome is the possibly right-censored time from selection to some later event. Again, SGL assumed monotonicity.

Monotonicity is a strong assumption – that one of the principal strata is empty. This assumption is never completely verifiable because participants are only assigned a single treatment. One could test the null hypothesis that the probability of selection in a treatment arm (eg, vaccine) is less than or equal to the probability of selection in the other arm (eg, placebo); if this null hypothesis is rejected (possibly within a subgroup defined by baseline covariates) then monotonicity is rejected. However, the rate of selection could be lower for treatment in all subgroups, yet monotonicity may still be violated. In the context of HIV vaccine trials, results from a recent trial also make it more difficult to assume monotonicity in future vaccine studies, as the risk of HIV-infection appeared higher in the vaccine arm than placebo within certain study subgroups (Buchbinder et al. 2008). Therefore, it is important to have methods for relaxing monotonicity.

Large sample bounds without assuming monotonicity were developed by Jemai (2005) and Zhang and Rubin (2003) for the average causal effect in the principal stratum of those individuals who would have been selected regardless of treatment assignment. These bounds are important, but represent extremes which are typically not plausible; and, as shown by Jemai (2005), in many cases these bounds are non-informative, as they are simply equivalent to plus or minus the difference between the minimum and maximum of the support of the outcome of interest. Jemai and Rotnitzky developed sensitivity analysis methods which relax the monotonicity assumption (Jemai 2005). These methods were applied by Shepherd, Redman, and Ankerst (2008) to an analysis investigating the effect of treatment (finasteride) on the severity of prostate cancer among men who would have had cancer regardless of treatment assignment.

In this manuscript, we develop methods that relax monotonicity and estimate the effect of treatment on a right-censored outcome in a subset of individuals who would be selected for inclusion regardless of their treatment assignment. These methods are an extension of Jemai (2005) (extending to a possibly right-censored outcome) and SGL (extending to relax monotonicity). In section 2 we specify our notation and assumptions. Section 3 addresses

estimation and inference, and section 4 considers specific scenarios which may arise in practice where either the number of selection events is fixed by the trial design or the selection event status of some participants is unknown because of censoring. Section 5 contains a simulation study to investigate finite sample behavior, section 6 applies these methods to data from an HIV vaccine trial, and section 7 discusses results.

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 one of two treatments, $Z_i = 0$ or $Z_i = 1$, for $i = 1, \dots, N$. Define S_i as a post-randomization selection variable where $S_i = 1$ implies an individual is selected, otherwise $S_i = 0$. For instance, in the HIV vaccine trial example, $S_i = 1$ if infected, $S_i = 0$ otherwise. Let T_i be the time from selection until some event for subject i , and define C_i as the time from selection until censoring. If $S_i = 1$, we observe $Y_i = \min(T_i, C_i)$ and $\delta_i = I(Y_i = T_i)$. Notice that T_i, C_i, Y_i , and δ_i only exist if $S_i = 1$; otherwise they are assigned the value $*$.

To define the estimand of interest, we use potential outcomes/counterfactuals (Neyman 1923, Rubin 1978, Robins 1986). Specifically, define $S_i(0)$ to be the selection indicator if, possibly contrary to fact, $Z_i = 0$. Define $S_i(1)$ to be the selection indicator if $Z_i = 1$. Similarly, define $T_i(0)$ to be the time-to-event outcome if $Z_i = 0$ and $T_i(1)$ to be the time-to-event outcome if $Z_i = 1$. The potential outcomes $C_i(z), Y_i(z)$, and $\delta_i(z)$ are similarly defined for $z = 0, 1$. This notation implicitly assumes that the potential outcomes of each trial participant are not influenced by the treatment assignments of other participants, an assumption known as no interference (Cox, 1958). If $S_i(z) = 0$, we define $T_i(z) = Y_i(z) = C_i(z) = \delta_i(z) = *$.

Assuming the study participants make up a random sample from a large population of interest, the outcomes $(Z_i, S_i(0), S_i(1), T_i(0), T_i(1), C_i(0), C_i(1))$, $i = 1, \dots, N$, are i.i.d. copies of a random vector $(Z, S(0), S(1), T(0), T(1), C(0), C(1))$, and similarly the observed data $O_i = (Z_i, S_i, Y_i, \delta_i)$, $i = 1, \dots, N$, are i.i.d. copies of $O = (Z, S, Y, \delta)$. We assume consistency, that $O_i(z) = O_i$ if $Z_i = z$ for all i . The assumptions of no interference and consistency constitute the stable unit treatment value assumption (SUTVA) (Rubin 1978). Random assignment of treatment ensures that

$$(S(0), S(1), T(0), T(1), C(0), C(1)) \perp\!\!\!\perp Z. \quad (1)$$

Here, for random variables A, B and C , $A \perp\!\!\!\perp B \mid C$ indicates conditional independence of A and B given C . We will also assume independent censoring:

$$C_i(z) \perp\!\!\!\perp T_i(z) \mid S_i(z) = 1 \text{ for } z=0, 1. \quad (2)$$

The four basic principal strata (Frangakis and Rubin 2002) can be defined in terms of the counterfactual pair $(S(0), S(1))$: (0,0), (0,1), (1,0), (1,1). For a subject i who is in the principal stratum defined by $(S(0), S(1)) = (1, 1)$, a causal effect on his/her time-to-event outcome is some measure of discrepancy between $T_i(0)$ and $T_i(1)$. Following SGL, we focus on the “survival” causal effect in the stratum $(S(0), S(1)) = (1, 1)$:

$$\begin{aligned} SCE(t) &= P(T_i(0) \leq t | S_i(0) = S_i(1) = 1) - P(T_i(1) \leq t | S_i(0) = S_i(1) = 1) \\ &\equiv F_0^{11}(t) - F_1^{11}(t). \end{aligned}$$

In order to estimate $SCE(t)$, in earlier work we made the following additional assumptions:

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

B.1: $P(S(1) = 1 | S(0) = 1, T(0)) = w_0(T(0); \beta_0)$, where $w_0(t; \beta_0) = G_0\{\alpha_0 + h_0(t; \beta_0)\}$, β_0 is fixed and known, $G_0(\cdot)$ is a known cdf, α_0 is an unknown parameter, and for each β_0 , $h_0(t; \beta_0)$ is a known function of t .

A.1, together with SUTVA, (1), and (2), identify $F_1^{11}(t) = F_1(t) \equiv P(T_i \leq t | Z_i = 1, S_i = 1)$. $F_0^{11}(t)$ is identified by further assuming B.1. The parameter β_0 is unknown and not identified by the observed data; it is a sensitivity parameter, treated as fixed and known in a single analysis and then varied over a plausible range of values to form a sensitivity analysis (Scharfstein, Rotnitzky, and Robins 1999; GBH). SGL specified $w_0(t; \beta_0)$ with the expit function,

$$w_0(t; \beta_0) = [1 + \exp\{-\alpha_0 - \beta_0 \min(t, \tau)\}]^{-1}. \quad (3)$$

where τ is some number less than or equal to the maximum length of follow-up. The expit function form for $w_0(\cdot)$ was originally used by GBH. With $w_0(\cdot)$ specified with the expit function, $\exp(\beta_0)$ has an odds ratio interpretation. Once β_0 is fixed, α_0 is identified by recognizing that $\int w_0(t; \beta_0) dF_0(t) = Pr(S = 1 | Z = 1) / Pr(S = 1 | Z = 0)$. SGL slightly modified $w_0(\cdot)$ (included $\min(t, \tau)$ as opposed to simply t) in order to make $w_0(\cdot)$ constant for $t > \tau$. With time-to-event data, follow-up is typically restricted so that $F_0(t)$ cannot be estimated nonparametrically for large t ; hence, SGL's modification permitted estimation of α_0 without assuming a parametric distribution for $F_0(t)$.

In this manuscript we relax assumption A.1 following an approach originally proposed by Jemai 2005. Instead of assuming A.1, we will assume B.1 and the following:

B.2: $P(S(0) = 1 | S(1) = 1, T(1)) = w_1(T(1); \beta_1)$, where $w_1(t; \beta_1) = G_1\{\alpha_1 + h_1(t; \beta_1)\}$, β_1 is fixed and known, $G_1(\cdot)$ is a known cdf, α_1 is an unknown parameter, and for each β_1 , $h_1(t; \beta_1)$ is a known function of t .

B.3: $\psi = \log \left(\frac{P(S(0) = 0, S(1) = 0) P(S(0) = 1, S(1) = 1)}{P(S(0) = 1, S(1) = 0) P(S(0) = 0, S(1) = 1)} \right)$, where ψ is fixed and known.

Assumption B.2 is analogous to B.1, and similar to (3), we will model $w_1(\cdot)$ with an expit function,

$$w_1(t; \beta_1) = [1 + \exp\{-\alpha_1 - \beta_1 \min(t, \tau)\}]^{-1}, \quad (4)$$

treating β_1 as a sensitivity parameter. Assumption B.3 specifies the joint distribution of $(S(0), S(1))$. The parameter ψ is not known, but will be treated as a sensitivity parameter and varied over a range of plausible values. ψ can be interpreted as the log odds ratio of selection under randomization $Z = 1$ for those selected versus those not selected under randomization $Z = 0$.

For ease of interpretation and presentation, we will additionally define the parameters $p_z = P(S(z) = 1)$, $\pi = P((S(0), S(1)) = (1, 1))$, and $\phi = P(S(0) = 1 | S(1) = 1)$, which will be helpful later in the paper. Instead of specifying the log-odds ratio sensitivity parameter ψ , one could specify π or ϕ . Specifying any of these (among other) parameters, together with the marginal probabilities p_0 and p_1 (which are identifiable from the observed data), specify the joint distribution of $(S(0), S(1))$. Therefore, assumption B.3 could be re-written in terms of π or ϕ . It should be noted that π is restricted to be between $\max(0, p_0 + p_1 - 1)$ and $\min(p_0, p_1)$, and ϕ is restricted to be between $\max(0, p_0 + p_1 - 1)/p_1$ and $\min(p_0, p_1)/p_1$, whereas ψ can take any value in $[-\infty, \infty]$.

Figure 1 demonstrates implications of certain sensitivity parameters with $p_1 < p_0$. Under monotonicity, $\psi = \infty$ (or equivalently $\pi = p_1$ and $\phi = 1$), and therefore the choice of β_1 is irrelevant as $S(0) = 1$ for all with $S(1) = 1$ (Figure 1, row 1). By relaxing monotonicity, $\psi < \infty$ (equivalently $\pi < p_1$ and $\phi < 1$) and it is unknown which subjects with $S(1) = 1$ are in the principal stratum defined by $(S(0), S(1)) = (1, 1)$, hence we specify β_1 (Figure 1, rows 2 and 3). Different choices of (ψ, β_0, β_1) lead to different values for $SCE(t=2)$. $F_z^{11}(t=2)$ for $z = 0, 1$ can be interpreted from Figure 1 as the proportion of the shaded curve less than or equal to $t = 2$. $SCE(2)$ is simply the difference between these cdfs.

Bounds for $SCE(t)$ are achieved by setting π to its minimum possible value and (β_0, β_1) to $(-\infty, \infty)$ and $(\infty, -\infty)$. The bounds for $SCE(t)$ can also be achieved at less extreme values of the sensitivity parameters (eg, Figure 1, row 4). If $p_0 + p_1 < 1$ then the bounds for $SCE(t)$ are the uninformative values -1 and 1 (Jemai 2005).

3. Estimation

Under B.1-B.3,

$$F_z^{11}(t_0) = \frac{\int_0^{t_0} w_z(t; \alpha_z, \beta_z) dF_z(t)}{\int w_z(t; \alpha_z, \beta_z) dF_z(t)}, \quad (5)$$

where $\int$ with unspecified limits of integration denotes $\int_0^\infty$. Note that for $z = 0, 1$,

$$\begin{aligned} \int w_z(t; \alpha_z, \beta_z) dF_z(t) &= \int P(S(1-z)=1 | S(z)=1, T(z)=t) P(T(z)=t | S(z)=1) ds \\ &= P(S(1-z)=1 | S(z)=1) \\ &= P(S(0)=S(1)=1) / P(S(z)=1) \\ &= \pi / P(S(z)=1) \\ &= \pi / P(S=1 | Z=z). \end{aligned} \quad (6)$$

We use this result to estimate α_z once the sensitivity parameters β_z and π have been specified. Specifically, $P(S=1 | Z=z)$ is consistently estimated by $\hat{p}_z = n_z / N_z$, where N_z and n_z are the number of subjects randomized and selected ($S_i = 1$), respectively, with $Z_i = z$. For given π, β_0, β_1 , we can then estimate α_z using (6), by plugging in the Kaplan-Meier estimates for $F_z(\cdot)$ and replacing $P(S=1 | Z=z)$ with $\hat{p}_z$. $F_z^{11}(t)$ can then be estimated by plugging in the estimate for α_z and the Kaplan-Meier estimates for $F_z(\cdot)$ into (5). The estimate for $SCE(t)$ then follows as the difference between the estimates of $F_0^{11}(t)$ and $F_1^{11}(t)$.

It should be recognized again that π is constrained to be within $[\max(0, p_0 + p_1 - 1), \min(p_0, p_1)]$ and that this estimation procedure will have problems if π is outside the range $[\max(0,$

$\hat{p}_0 + \hat{p}_1 - 1$, $\min(\hat{p}_0, \hat{p}_1)$], where $\hat{p}_z = n_z/N_z$. For example, if $\pi > n_1/N_1$ then since $0 \leq \int w_1(t; \alpha_1, \beta_1) dF_1(t) \leq 1$, there is no value of α_1 that will solve $\int w_1(t; \alpha_1, \beta_1) d\hat{F}_1(t) = \pi N_1/n_1$. An ad hoc fix in this situation is to set $\pi = n_1/N_1$ and perform the analysis. Alternatively, one could choose a range for plausible values of π (or equivalently ϕ) based on the estimated values $\hat{p}_z$. However, this is a little unsettling as π is treated as fixed and known in the analysis. We will follow Jemai (2005) and avoid this problem by specifying the unrestrained parameter ψ and then recognizing that

$$\pi(\psi, p_0, p_1) = \begin{cases} p_0 p_1 & \text{if } \psi = 0 \\ \frac{1 + (p_1 + p_0)(e^\psi - 1) - \sqrt{(1 + (p_1 + p_0)(e^\psi - 1))^2 - 4(e^\psi - 1)p_0 p_1 e^\psi}}{2(e^\psi - 1)} & \text{if } \psi \neq 0. \end{cases}$$

Therefore, for a given ψ , $\pi(\psi, p_0, p_1)$ will be estimated as $\pi(\psi, \hat{p}_0, \hat{p}_1)$. If results are presented using the sensitivity parameter ϕ instead of ψ (as in the Section 6), then it should be recognized that ϕ must satisfy $\pi(\psi, \hat{p}_0, \hat{p}_1)/\hat{p}_1$.

Our estimator for $SCE(t)$ is consistent and asymptotically normal, following the results of SGL. Confidence intervals for $SCE(t)$ can therefore be computed using bootstrap techniques, sampling with replacement N records from O_i (described in more detail in section 5).

4. Additional Considerations

4.1 Fixed number of selection events

In some studies, the trial continues until a pre-specified number of total events are accrued (i.e., until $\sum S_i = n$, where n is fixed before the trial begins). For example, some HIV vaccine trials continue until a certain number of infections are observed. In this situation, estimation of $SCE(t)$ is identical, although variance estimates may need to be adjusted. A slight variation of the bootstrap confidence intervals is sufficient: Instead of sampling N records with replacement from O_i , sample with replacement from O_i until there are $n = n_0 + n_1$ records with the selection event.

4.2 Selection status unknown for some participants

In many studies, the selection event S is a dichotomization of a failure-time outcome. For example, in the HIV vaccine example, S is the indicator of HIV-infection during follow-up. If follow-up varies between participants, one may want to define selection as the occurrence of the selection event (e.g., infection) by time k after randomization. In the HIV vaccine example, one could consider the effect of vaccination on time from infection diagnosis to initiating ART among the subgroup of subjects who would have been infected by time k regardless of treatment assignment. Some subjects may have been censored before time k so their event status at time k is unknown. However, the estimation procedure given in section 3 can be easily extended to account for missing S due to non-informative censoring.

Specifically, re-define S to be the indicator that the event occurred by time k . Those who did not have the event by time k are assigned $S = 0$ and any subsequent events are ignored, and those who were not followed until time k are censored at the time of their last visit. Assuming the time until censoring is independent from the time until the selection event, $P(S = 1|Z = z)$ is consistently estimated with a Kaplan-Meier estimate at time k . Therefore, estimation proceeds as outlined in section 3 except $\hat{p}_z$ are now Kaplan-Meier estimates. Note that an individual who is not selected by time k contributes no post-selection data to the analysis (i.e., Y_i and δ_i are not included if $S_i = 0$, even though this information may be

available for a different choice of k). Confidence intervals for this approach can also be computed using the bootstrap, sampling N vectors with replacement from O_i .

5. Simulations

5.1 Primary Simulation Scenarios

To evaluate the performance of our estimators we conducted several simulation experiments. Each simulation generated N data vectors according to the following steps:

1. The first $N/2$ vectors were set at $Z = 0$; the second $N/2$ at $Z = 1$.
2. For those with $Z = z$, S was drawn from a Bernoulli (p_z) distribution with $p_z = 0.10$ and 0.05 for $z = 0$ and 1 , respectively.
3. For those with $Z = z$ and $S = 1$, T was generated from the probability density function

$$f_z(t) = \frac{\pi(\psi, p_0, p_1) \left(\frac{a}{b}\right) \left(\frac{t}{b}\right)^{a-1} \exp\left\{-\left(\frac{t}{b}\right)^a\right\}}{[1 + \exp\{-\alpha_z - \beta_z \min(t, \tau)\}]^{-1} p_z},$$

where $a = 0.4$, $b = 9$, $\psi = 4.08$, $\tau = 24$, $\beta_0 = -0.1$, $\alpha_0 = 0.981$, $\beta_1 = -0.1$, $\alpha_1 = 2.773$.

The choices for α_z were determined by solving the equation $\int_0^\infty f_z(t) dt = 1$ by numerical integration given the other parameter values. The choice of $\psi = 4.08$ corresponds to $\pi(\psi, p_0, p_1) = 0.04$. T was generated from $f_z(t)$ using rejection sampling (von Neumann 1951; Liu 2001). Specifically, t was drawn from a Weibull distribution with shape and scale parameters equal to a and b , respectively. We then computed

$$r(t) = \frac{1 + \exp\{-\alpha_z - \beta_z \min(t, \tau)\}}{1 + \exp\{-\alpha_z - \beta_z \tau I_{(\beta_z < 0)}\}},$$

where $I_{(\cdot)}$ is the indicator function, and accepted $T = t$ if a Bernoulli trial with probability of success $= r(t)$ was a success, otherwise t was discarded. This process was repeated until we accepted some value of t .

4. For those with $S = 1$, C_1 was generated from a Uniform (0, 100) distribution. Then C was set as $\min(\tau, C_1)$.
5. For all realizations with $S = 1$, Y was chosen as $\min(C, T)$ and $\delta = I_{(Y=T)}$.

It can be shown that these steps result in data consistent with SUTVA, (1), (2), B.1, B.2, B.3, and models (3) and (4), where given membership in the principal stratum defined by $S(0) = S(1) = 1$, the distribution of $T(0)$ and $T(1)$ are identical and follow a Weibull distribution with shape and scale parameters a and b , respectively. Therefore $SCE(t) = 0$ for all t .

For each data generating scenario, we generated 1000 datasets and estimated $SCE(t)$ at $t = 24$ months with β_0 , β_1 , and ψ properly specified. For each estimate, 95% confidence intervals were computed using 1000 bootstrap replications: Wald-based confidence intervals ($\pm 1.96 \times$ standard deviation of 1000 bootstrap estimates of $SCE(24)$), and percentile-based confidence intervals (2.5th and 97.5th percentiles of the 1000 bootstrap estimates).

Table 1 summarizes our method's behavior when the sensitivity parameters were properly specified in 1000 simulation replications for $N=1000, 2000, 5000$, and 10000 . Corresponding expected values for n_0 and n_1 are given in the table; corresponding values for the expected number of events are $E(\sum (1 - Z_i) S_i \delta_i) \approx Np_0/4$ and $E(\sum Z_i S_i \delta_i) \approx Np_1/3$, which when $N = 1000$ corresponds to approximately 25 and 17 expected events. The coverage probability was near 0.95 for all N . Bias was consistently negative in small samples, though quite small. We suspect that the known bias of Kaplan-Meier estimators in small-samples is leading to the slight bias seen here. In small samples, 1-sided Type I error rates for testing $H_0: SCE(t) \leq 0$ were inflated, corresponding to the small sample bias. (One-sided Type I error rates for testing $H_0: SCE(t) \geq 0$ were correspondingly less than 0.05 in small samples; data not shown.)

We also performed simulations for $N=5000$, varying β_0 and β_1 using all possible 169 combinations of $\{-0.3, -0.25, -0.2, \dots, 0.25, 0.3\}$ with the sensitivity parameters correctly specified. (α_z changed accordingly, all other parameters remained the same.) Simulation results are shown in Figure 2. Bias across all 169 simulations ranged from -0.012 to 0.006; those scenarios which resulted in absolute bias greater than 0.004 are shaded in the upper plot. Coverage ranged from 0.929 to 0.964 using bootstrap percentile confidence intervals and 0.919 to 0.966 using Wald-based bootstrap confidence intervals. We found no apparent patterns in bias or coverage, except that simulations with $\beta_1 = -0.3$ or -0.25 and $\beta_0 \leq 0$ tended to have lower coverage when using Wald-based intervals (Figure 2 lower plot).

It should be noted that these simulations were set-up to be similar to the HIV vaccine trial, with T in terms of months. Therefore, for example, $\beta_0 = 0.3$ corresponds to an odds ratio of $\exp(0.3 \times 12) = \exp(3.6) \approx 36.6$ when interpreted in terms of a one-year change (which is how we will scale the odds ratios in section 6).

We performed an additional simulation investigating our method with $N = 1000$ but higher rates of selection: $p_0 = 0.90, p_1 = 0.95, \pi = 0.89, \beta_0 = \beta_1 = -0.1, \alpha_0 = 5.87$, and $\alpha_1 = 4.08$. Coverage of 95% confidence intervals was 0.955 and 0.959 for percentile and Wald-based intervals, respectively, and bias was small, -0.0008 .

5.2 Additional Simulation Scenarios

We performed additional simulations under the scenario where a trial is continued until a fixed number of selection events occur (section 4.1). To do this, data were generated in a manner identical to that described above except N was no longer a fixed number and we alternately sampled S from Bernoulli(p_0) and Bernoulli(p_1) distributions until $n = 75, 150, 375$, and 750 . Simulation results are given in the bottom half of Table 1 and are similar to those where the number of study participants were fixed.

Finally, we performed simulations under the scenario where the selection status was unknown for some participants due to censoring (section 4.2). To do this, data were generated in a manner identical to that described above except step 2 was replaced with the following:

- 2a. For those with $Z = z$, the time from randomization to selection, U , was drawn from an exponential distribution with rate parameter λ_z , with $\lambda_z = -\log(1 - p_z)/36$ and $p_z = 0.10$ and 0.05 for $z = 0$ and 1 , respectively. Hence, $P(U \leq 36|Z = z) = p_z$.
- 2b. The time from randomization to censoring, C^* , was generated from a Uniform(0,100) distribution.
- 2c. The observed pre-selection follow-up, V , was chosen as $\min(C^*, U)$ and the indicator of selection, $S^* = I_{(V = U)}$.

2d. S was set as equal to 1 if $S^* = 1$ and $V \leq 36$; $S = 0$ if $V > 36$ (regardless of whether $S^* = 0$ or 1); and S was missing if $S^* = 0$ and $V \leq 36$.

This data generating scenario mimics a study where selection is defined as the selection event occurring within 36 months of randomization, and where some participants are not followed 36 months so their selection event status is unknown. In these simulations, data were generated with $N=610, 1220, 3050$, and 6100 , resulting in approximately $n=75, 150, 375$, and 750 observed selection events, similar to the other simulations. Results of these simulations are given in the bottom third of Table 1 and are similar to the earlier simulations.

6. Example

We illustrate our methods using data from a randomized, double-blind, placebo-controlled Phase III trial of AIDSVAX B/B conducted from 1998-2003. 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; the estimated vaccine efficacy (defined as one minus the relative risk of infection) was 0.048, corresponding to 6.7% (241/3598) and 7.0% (127/1805) infected in the vaccine and placebo arms, respectively. However, interaction tests suggested that the vaccine might partially prevent infection for non-whites. Among non-whites vaccine efficacy was estimated as 0.47, corresponding to 5.0% (30/604) infected in the vaccine arm and 9.4% (29/310) infected under placebo. Detailed results are found in Flynn et al. (2005). Here we compare the time from infection diagnosis to the initiation of antiretroviral therapy (ART) between the vaccine and placebo arms among non-whites who would have been infected regardless of randomization assignment. Specifically, we perform sensitivity analyses to test the hypothesis $H_0: SCE(t) = 0$ for $t = 2$ years; $Z = 0/1$ corresponds to randomization to placebo/vaccine, S is the indicator of infection, and $T =$ time from infection diagnosis to initiating ART. A vaccine effect to delay ART may be beneficial because it delays exposure to drug toxicities, drug resistance, and the depletion of future therapy options. SGL also estimated $SCE(t)$ at $t = 2$ years but assumed monotonicity (reproduced here as Figure 3). Here we repeat the analysis relaxing the assumption of monotonicity.

Figure 4 shows a sensitivity analysis of $SCE(t = 2)$, varying ϕ , β_0 , and β_1 . The range for β_0 and β_1 is $[-3, 3]$, corresponding to odds ratios of $[0.05, 20]$ per year. Similar to SGL, $\tau = 2$ years, corresponding to approximately the maximum length of follow-up post-infection diagnosis. Contours represent the estimated $SCE(2)$ at a given ϕ , β_0 , and β_1 . Shaded regions correspond to those sensitivity parameters where the Wald-based 95% bootstrap confidence intervals for $SCE(2)$ do not contain 0. Estimates in the dark-shaded regions imply that among those who would have been infected regardless of treatment assignment, vaccination lowered the probability of starting ART within 2 years of infection diagnosis (was beneficial), whereas estimates in the light-shaded regions imply that vaccination increased the probability of starting ART by 2 years (was harmful). The sensitivity parameter ϕ is defined as the probability of infection if randomized to placebo given infection when randomized to vaccine. ϕ varies from 0.99, corresponding to a minor departure from monotonicity, to $\phi = \hat{p}_0 = 0.094$, implying that whether someone gets infected in the vaccine arm is independent of whether they would have been infected if randomized to placebo. This independence assumption seems unlikely; even more unlikely is that $S(0)$ and $S(1)$ are negatively correlated, so plots assuming $\phi < \hat{p}_0$ are not shown.

As expected, results when $\phi = 0.99$ closely correspond to the analysis under monotonicity ($\phi = 1$). In both cases, $SCE(2) = 0$ is rejected for $\exp(\beta_0) > 1$. Under monotonicity, the choice of β_1 is irrelevant; β_1 must be specified when $\pi = 0.99$, but its choice has little impact on

results. As we get farther away from monotonicity (ie, smaller values of ϕ), the choice of β_1 has greater impact. Furthermore as ϕ gets smaller, the range of values for $SCE(2)$ increases. For $\phi = \hat{p}_0$, the estimate of $SCE(2)$ ranges from -0.99 to 0.78 , for β_0 and β_1 in $[-3, 3]$. If we were to estimate $SCE(2)$ over the entire range of possible sensitivity parameter values, we would achieve the lower and upper bounds, -1 and 1 .

Based on these sensitivity analyses, it is very unlikely that the vaccine is having a beneficial effect by delaying the initiation of ART among those who would have been infected no matter their treatment assignment. $SCE(2)$ is only statistically greater than zero if there are extreme violations of monotonicity (e.g., $<50\%$ of those infected in the vaccine arm would have been infected if randomized to placebo) and if OR_0 and OR_1 are quite different from 1 and in opposite directions. If anything, it is more likely the vaccine is having a harmful effect as under more plausible scenarios (ϕ close to one and OR_0 slightly greater than one), there is evidence that $SCE(2)$ is less than zero. However, based on our past elicitation experiences with subject matter experts (SGL; Shepherd, Gilbert, and Mehrotra 2006), those with worse post-infection outcomes under placebo were thought to be those who would have been more likely infected if randomized to vaccine, implying that $OR_0 \leq 1$, and for these values the 95% confidence intervals for $SCE(2)$ include zero.

Although we have focused our analysis on the difference between the probability of ART initiation 2 years after infection diagnosis ($SCE(2)$), our methods allow estimation of the distribution of the time until ART initiation under vaccine or placebo in the always selected principal stratum, and hence estimates of $SCE(t)$ for any t . Figure 5 shows how estimates of these distributions and $SCE(t)$ change with respect to a few limited choices of the sensitivity parameters. It is challenging to present estimates and confidence intervals from a 3-parameter sensitivity analysis at multiple time points. Such analyses may best be presented with an animation: The supplementary appendix contains an animated sensitivity analysis which shows the estimated $SCE(t)$ in the non-white cohort of the VaxGen trial as a function of the sensitivity parameters for each month post-infection (e.g., for $t = 1/12, 2/12, \dots, 24/12$).

Finally, we also performed analyses to estimate $SCE(2)$ where the principal stratum of interest was defined as infection by time k post-randomization regardless of treatment assignment. For $k = 36$ months, estimates of $SCE(2)$ were very similar to those presented in Figure 4. That is because 97% of our observed infections occurred by time $k = 36$ months, and estimates of $\hat{p}_z$ were quite similar to those where we ignored the length of follow-up. When $k = 24$ months, results differed a little more as only 66% of our observed infections had occurred by this time. These analyses are presented in the supplementary appendix.

7. Discussion

We have described a sensitivity analysis approach for assessing treatment effects on a time-to-event outcome measured in a subset selected after randomization. Our approach extends earlier work by relaxing the monotonicity assumption – that one of the principal strata was empty – and including scenarios where selection status may not be known for some subjects because of censoring. We applied our method to data from an HIV vaccine trial.

Another potential application of these methods is the assessment of a randomized treatment effect in the “per-protocol” sub-population that received the full treatment. For example, it is standard in vaccine efficacy trials to assess the effect of vaccination on time-to-clinically significant infection in the subgroup of subjects who receive all of the protocol-scheduled injections (Horne et al., 2000). However, virtually always in the published literature the analysis assesses the treatment effect in the subgroup of subjects observed to be in the per-

protocol population, and hence is susceptible to selection bias. Another relevant example comes from a randomized trial to assess the effect of breast- versus formula-feeding on the composite outcome of HIV infection or death among infants who are alive and HIV-free shortly after birth (Thior et al., 2006). The purpose of such an analysis is to isolate the effect of breastfeeding versus formula feeding on HIV-free survival, as many infants acquire HIV or die during, before, or shortly after birth. The methods in this article provide approaches for sensitivity analyses to estimate the causal treatment effect in both of these studies.

Our example analysis illustrated the application of the method to a situation where selection (HIV-infection) was rare and the selection rate for one arm was nearly twice that of the other. In these scenarios, results are highly variable across different sensitivity parameters. In contrast, when selection is common (e.g., receiving all protocol-scheduled injections in the example mentioned above), estimates will be less variable across sensitivity parameter values, and relaxing monotonicity will have less impact on the results.

Methods based on principal stratification such as those presented here, are helpful for better understanding causal mechanisms. However, these types of methods are not without limitations. Their utility for decision-making is limited as 1) principal strata membership is almost always unknown, and 2) most decisions regarding treatment use need to take into account both the effect of the treatment on selection as well as the post-selection outcome (Joffe, Small and Hsu, 2007). For example, licensure of an HIV vaccine will be based on the vaccine's ability to prevent infection, and not just its ability to improve health among those who would get infected regardless of treatment assignment. In addition, principal stratification methods require estimation within discrete strata which may be an oversimplification of the underlying causal mechanisms. For example, infection was treated as a dichotomous outcome, yet a vaccine may affect the timing of infection which then may affect the post-infection outcome. It is conceivable that a vaccine prevents infections from exposure to weak viruses early in a trial, thereby allowing some participants to be infected later in the trial with a more virulent virus. Therefore, although such a participant would be protected by the vaccine, this person would still be in the always infected principal stratum and would have a worse post-infection outcome when assigned vaccine than when assigned placebo (as discussed by Robins, Rotnitzky and Vansteelandt (2007) in a different context). One could define principal strata based on infection by a specific time from randomization, as presented in this manuscript. However, to fully address this problem one would need to consider many different principal strata which may not be feasible given data sparseness. An associate editor suggested redefining Y as the time from randomization until the event. In the HIV vaccine context, $SCE(t)$ would then be a contrast of the probabilities of starting ART t months *post-randomization* among those who would have been infected by month k regardless of treatment assignment. The supplementary appendix contains the results of such an analysis applied to the VaxGen data. This analysis partially addresses the concern mentioned above because it includes both time to infection and time from infection to starting ART in the outcome. However, because the outcome is comprised of two events, this approach may provide less mechanistic insight. In summary, it should be recognized that even in the principal stratification framework, causal mechanisms may not be readily apparent.

One can also imagine similar scenarios that could lead to violations of the independent censoring assumption (2). For example, those infected earlier with longer potential lengths of post-infection follow-up may be those with weaker immune systems susceptible to more rapid disease progression. This could motivate one to focus on principal strata defined using shorter amounts of time k from randomization until selection. However, in the HIV vaccine trial example, infected subjects were followed a fixed amount of time after infection (2 years), so this potential problem is minimal.

Another challenge to our method is simultaneously specifying and interpreting three sensitivity parameters. Eliciting counterfactual sensitivity parameters can be challenging, and final results, with their caveats and potentially different conclusions at different sensitivity parameter values, are usually not amenable to concise statements. One could assign distributions (priors) to our sensitivity parameters and integrate over these parameters to arrive at a single answer. However, because the answer depends on the chosen prior distributions, this approach does not obviate the need for a sensitivity analysis; rather it would be relevant to repeat the analysis for a set of plausible priors, and this set would need to be elicited. We have discussed elicitation and interpretation of sensitivity parameters elsewhere (Shepherd, Gilbert, and Mehrotra 2006).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported in part by NIH grants R01DA023879 and 2R37AI054165. The authors would like to thank the editors and 2 referees for their helpful suggestions.

References

- Buchbinder SP, Mehrotra DV, Duerr A, et al. Efficacy assessment of a cell-mediated immunity HIV-1 vaccine (the Step Study): a double-blind, randomised, placebo-controlled, test-of-concept trial. *Lancet*. 2008; 372:1881–93. [PubMed: 19012954]
- Cox, DR. *Planning of Experiments*. New York: Wiley; 1958.
- Frangakis CE, Rubin DB. Principal stratification in causal inference. *Biometrics*. 2002; 58:21–29. [PubMed: 11890317]
- Gilbert PB, Bosch RJ, Hudgens MG. Sensitivity analysis for the assessment of causal vaccine effects on viral load in HIV vaccine trials. *Biometrics*. 2003; 59:531–541. [PubMed: 14601754]
- Graham BS. Clinical trials of HIV vaccines. *Annual Review of Medicine*. 2002; 53:207–221.
- Horne AD, Lachenbruch PA, Goldenthal KL. Intent-to-treat analysis and preventive vaccine efficacy. *Vaccine*. 2000; 19:319–326. [PubMed: 10930687]
- Hudgens MG, Hoering A, Self SG. On the analysis of viral load endpoints in HIV vaccine trials. *Statistics in Medicine*. 2003; 22:2281–2298. [PubMed: 12854093]
- Jemai, Y. Unpublished PhD dissertation under the supervision of A Rotnitzky. Harvard School of Public Health, Department of Biostatistics; 2005. Semiparametric methods for inferring treatment effects on outcomes defined only if a post-randomization event occurs.
- Joffe MM, Small D, Hsu CY. Defining and estimating intervention effects for groups that will develop an auxiliary outcome. *Statistical Science*. 2007; 22:74–97.
- Neyman J. On the application of probability theory to agricultural experiments: Essay on principles. Translated in *Statistical Science*. 1923; 5:465–480.
- Liu, JS. *Monte Carlo Strategies in Scientific Computing*. Springer-Verlag; New York: 2001.
- Mehrotra DV, Li X, Gilbert PB. A comparison of eight methods for the dual-endpoint evaluation of efficacy in a proof-of-concept HIV vaccine trial. *Biometrics*. 2006; 62:893–900. [PubMed: 16984333]
- Robins JM. A new approach to causal inference in mortality studies with sustained exposure periods - Application to control of the healthy worker survivor effect. *Mathematical Modeling*. 1986; 7:1393–1512.
- Robins J, Rotnitzky A, Vansteelandt S. Discussion of “Principal stratification designs to estimate input data missing due to death” by Frangakis CE, Rubin DB, An MW, and MacKenzie E. *Biometrics*. 2007; 63:650–651. [PubMed: 17824996]
- Rosenbaum PR. The consequences of adjustment for a concomitant variable that has been affected by the treatment. *Journal of the Royal Statistical Society, Series A*. 1984; 147:656–666.

- Rubin DB. Bayesian inference for causal effects: the role of randomization. *The Annals of Statistics*. 1978; 6:34–58.
- Scharfstein DO, Rotnitzky A, Robins JM. Adjusting for nonignorable drop-out using semiparametric nonresponse models. *Journal of the American Statistical Association*. 1999; 94:1096–1146.
- Shepherd BE, Redman MW, Ankerst DP. Does finasteride affect the severity of prostate cancer? A causal sensitivity analysis. *Journal of the American Statistical Association*. 2008; 484:1392–1404. [PubMed: 20526381]
- Shepherd BE, Gilbert PB, Lumley T. Sensitivity analyses comparing time-to-event outcomes only existing in a subset selected post-randomization. *Journal of the American Statistical Association*. 2007; 102:573–582. [PubMed: 19122791]
- Shepherd BE, Gilbert PB, Mehrotra DV. Eliciting a counterfactual sensitivity parameter. *The American Statistician*. 2007; 61:56–63.
- Thior I, Lockman S, Smeaton LM, et al. Breastfeeding plus infant zidovudine prophylaxis for 6 months vs formula feeding plus infant zidovudine for 1 month to reduce mother-to-child HIV transmission in Botswana: a randomized trial: the Mashi Study. *Journal of the American Medical Association*. 2006; 296:794–805. [PubMed: 16905785]
- von Neumann J. Various techniques used in connection with random digits. *National Bureau of Standards Applied Mathematics Series*. 1951; 12:36–38.
- Zhang JL, Rubin DB. Estimation of causal effects via principal stratification when some outcomes are truncated by ‘death’. *Journal of Educational and Behavioral Statistics*. 2003; 28:353–368.

Figure 1.

Illustration of sensitivity parameters. The outer curve represents the distribution of $T(z)|S(z) = 1$. The shaded curve represents the distribution of $T(z)|S(0) = S(1) = 1$. All curves were generated under (3)-(4) with $p_0 = 0.10$, $p_1 = 0.05$, $\tau = 3$, and $T(z)|S(z) = 1$ for $z = 0, 1$ from Weibull(1.5,1.5). $\phi \equiv P(S(0) = 1|S(1) = 1)$ and $SCE(2) \equiv F_0^{II}(2) - F_1^{II}(2)$. Row 1: Under assumption of monotonicity. Rows 2-3: Relaxing monotonicity. Row 4: Relaxing monotonicity and achieving the upper bound for $SCE(2)$.

Figure 2.

Simulation results for $\beta_0, \beta_1 \in \{-0.3, -0.25, \dots, 0.3\}$. Shaded regions for the lower two plots correspond to coverage outside the expected Monte Carlo error $0.95 \pm 1.96 \sqrt{0.96(1 - 0.95)/1000} = 0.9365$ and 0.9635 .

Figure 3. Sensitivity analysis of $SCE(2)$ for the non-white cohort of VaxGen's trial under monotonicity ($\phi=1$).

Figure 4.
Sensitivity analysis of $SCE(2)$ for the non-white cohort of VaxGen's Phase III trial.

Figure 5.

Sensitivity analysis for $F_0^{11}(t)$ (top), $F_1^{11}(t)$ (middle), and $SCE(t)$ with $\phi = 0.90$ in the non-white cohort of VaxGen's Phase III trial.

Table 1

Simulation Results ($\beta_0 = \beta_1 = -0.1$)

E(n0)		E(n1)	Bias	Coverage		1-Sided Type I Error	
				Percentiles	Wald	Percentiles	Wald
N fixed							
N							
1000	50	25	-0.0115	0.954	0.953	0.072	0.063
2000	100	50	-0.0030	0.941	0.947	0.068	0.067
5000	250	125	0.0004	0.937	0.935	0.064	0.058
10000	500	250	-0.0011	0.939	0.942	0.062	0.058
n fixed							
n							
75	50	25	-0.0091	0.947	0.948	0.070	0.065
150	100	50	-0.0041	0.937	0.946	0.069	0.062
375	250	125	-0.0006	0.950	0.950	0.047	0.047
750	500	250	-0.0025	0.939	0.944	0.062	0.064
selection status unknown for some participants							
N							
610	50	25	-0.0078	0.933	0.945	0.082	0.068
1220	100	50	0.0009	0.945	0.949	0.069	0.066
3050	250	125	-0.0023	0.952	0.955	0.051	0.054
6100	500	250	-0.0014	0.952	0.953	0.056	0.049