Eliciting a Counterfactual Sensitivity Parameter

Bryan E. Shepherd
Department of Biostatistics,
Vanderbilt University, Nashville, TN, 37232, USA
bryan.shepherd@vanderbilt.edu

Peter B. Gilbert
Department of Biostatistics,
University of Washington, Seattle, WA, 98195, USA

Devan V. Mehrotra
Merck Research Laboratories,
North Wales, PA, 19454, USA

Sensitivity analyses, wherein estimation is performed for each of a range of values of a sensitivity parameter, are of particular use in causal inference, where estimands often are identified because of untestable assumptions. Sensitivity parameters may have counterfactual interpretations, making their elicitation especially challenging. Here we describe our experience eliciting a counterfactual sensitivity parameter to be used in the analysis of an ongoing HIV vaccine trial. We include instructions given to 10 subject-matter experts, the chosen ranges of our 8 responders and some of their comments, their ranges applied to data from an earlier trial, and a brief discussion of some general issues regarding counterfactual elicitation.

KEYWORDS: Causal inference; principal stratification; elicitation; AIDS.

1. INTRODUCTION

This paper is motivated by the primary analysis for an ongoing HIV vaccine efficacy trial. The analysis for this trial requires elicitation of a sensitivity parameter in order to evaluate the robustness of inference to selection bias. In this and similar trials, high-risk HIV-negative volunteers are randomized to vaccine or placebo and then followed for a period of time during which some participants will become HIV infected. Among other things, researchers are interested in knowing the causal effect of vaccination on set-point viral load (a measurement of the concentration of virus in an infected person's plasma shortly after infection). The current wave of candidate HIV vaccines are specifically designed to lower (improve) set-point viral loads for those who become infected, perhaps making infected individuals more healthy and less likely to pass on the virus. Merck and the HIV Vaccine Trials Network (HVTN) are currently conducting a vaccine trial in which the primary outcomes are HIV infection and set-point HIV viral load (Mehrotra, Li, and Gilbert 2006). As is well-documented, an analysis comparing the vaccine and placebo arms that conditions on whether or not participants get infected could be subject to selection bias (Rosenbaum 1984).

Gilbert, Bosch, and Hudgens (2003) (GBH) proposed using a counterfactual sensitivity parameter to deal with this potential post-randomization selection bias. A sensitivity parameter is a quantity used to identify an estimand while allowing one to examine robustness to a gradation of identifying assumptions. Typically, a sensitivity parameter is assumed fixed and known to allow estimation, but varied over a range

of values in a sensitivity analysis, producing a range of estimates. Sensitivity parameters are of particular use in causal inference, where untestable assumptions are often needed to identify a causal estimand. Such sensitivity parameters may have a counterfactual interpretation. A counterfactual expression is of the form: if, contrary to fact, B had occurred, then C would have been the result. In reality B did not occur; therefore, the role of the sensitivity parameter is often to propose different possible outcomes C, had, contrary to fact, B occurred. The sensitivity parameter proposed by GBH considers people who got infected in the placebo arm and uses their viral load to assign odds of infection if, contrary to fact, they had been randomized to the vaccine arm (further described in Section 2).

The use of counterfactual sensitivity parameters has been proposed in many different settings. Leamer (1974) suggested using sensitivity parameters to represent bias in the coefficients of a regression model due to the exclusion of other variables from the model. In observational studies, sensitivity parameters have been used to assess the affect of unmeasured confounders (e.g., Rosenbaum and Rubin 1983; Rosenbaum 1987; Rosenbaum 1988; Greenland 1996; Copas and Li 1997; Greenland 2005). Sensitivity parameters have also been used to address non-ignorable missing data; for example, Little (1994) proposed using a sensitivity parameter in pattern-mixture models for incomplete data, and Scharfstein, Rotnitzky, and Robins (1999) suggested using sensitivity parameters to account for non-ignorable dropout. In all of these situations, estimands can only be identified after making an untestable assumption

(e.g., a complete model, no unmeasured confounders, ignorable missingness, etc.); the sensitivity parameter allows identification and evaluation under varying assumptions. These sensitivity parameters also have counterfactual interpretations. For example, if, contrary to fact, a person who had dropped out had continued to be followed, what is the probability they would have died before the end of the study? Or if, contrary to fact, an unobserved confounder had been measured, what would have been its influence on the odds of treatment and on the value of the outcome?

Subject-matter experts must be involved in the process of selecting plausible ranges for sensitivity parameters. However, sensitivity parameters may be difficult to elicit due to disconnects that often exist between statisticians and other scientists. Sometimes techniques or parameters are challenging to explain to other statisticians, let alone biologists with little statistical training. In many ways, eliciting a plausible range for a sensitivity parameter is similar to eliciting prior distributions; challenges of prior elicitation have been discussed, and progress has been made in developing techniques for overcoming elicitation difficulties (for example, see Chaloner 1996; Kadane and Wolfson 1998; O'Hagan 1998). However, many elicitation challenges are compounded when dealing with a counterfactual sensitivity parameter because of the difficulties explaining and interpreting unobservable counterfactuals. Indeed, one of the most common complaints against the use of counterfactual sensitivity parameters is that they are too hard to elicit from subject-matter experts.

In this paper, we describe our experience eliciting a plausible range for a coun-

terfactual sensitivity parameter from ten HIV vaccine experts.

2. MODELING THE CAUSAL EFFECT

One can remove selection bias and assess the causal effect of vaccination on viral load by comparing viral loads between individuals in the vaccine and placebo arms who would have been infected regardless of treatment assignment (Robins 1986; Rubin 2000; Frangakis and Rubin 2002). These individuals have been referred to as belonging to the always infected principal stratum. One causal estimand of interest is the average causal effect (ACE), the mean difference in viral loads conditional on being infected under either treatment assignment,

$$ACE \equiv E(Y(1) - Y(0)|S(0) = S(1) = 1). \tag{1}$$

Here, $S(z)$ is an indicator of being infected if randomized to treatment z; $z = 0, 1$ implies randomization to the placebo, vaccine arms, respectively; and $Y(z)$ is the post-infection viral load if randomized to treatment z, where $Y(z)$ does not exist (is set equal to $*$) if $S(z) = 0$ (uninfected). The variables $S(0), S(1), Y(0)$, and $Y(1)$ are referred to as potential outcomes / counterfactuals, and are defined for all subjects even though for any given subject $(S(z), Y(z))$ will be observed for only $z = 0$ or $z = 1$ but not both (Holland 1986).

GBH proposed a method for estimating ACE, based on the following assumptions:

A.1 Stable Unit Treatment Value Assumption: that the potential outcomes of

each trial participant are not influenced by the treatment assignments of other participants (Cox 1958; Rubin 1978).

A.2 Randomization: $(S(0), S(1), Y(0), Y(1))$ is independent of Z.

A.3 Monotonicity: $S(1) \leq S(0)$, i.e., that everyone infected in the vaccine arm would have been infected if assigned placebo.

A.4 $P(S(1) = 1|S(0) = 1, Y(0) = y) = \{1 + exp(-\alpha - \beta y)\}^{-1}$, where β is a known constant, and α is an unknown parameter to be estimated.

Assumptions A.1-A.3 are thought to be reasonable in this HIV vaccine trial setting (GBH, Shepherd et al. 2006), and will be assumed throughout. A.1-A.3 alone do not identify ACE, whereas A.1-A.4 do; A.4 supposes a parametric form for the counterfactual probability of infection if randomized to the vaccine arm given infection when randomized to the placebo arm, the placebo viral load, and a value for β. β can be interpreted as the log-odds ratio of infection if randomized to the vaccine arm for a 1-unit increase in placebo viral load given infection in the placebo arm. If $\beta > 0$ then those with higher viral loads in the placebo arm are more likely to be those who would have been infected regardless of treatment assignment; hence, as β increases ACE decreases. In actuality, β is unknown and unidentifiable, which led GBH to advocate performing a sensitivity analysis in which β is varied over a plausible range of values. It is therefore imperative to come up with a scientifically plausible range of values for the counterfactual sensitivity parameter β. Can such a range be obtained? What are plausible ranges for β in the Merck/HVTN trial?

3. INFORMAL EMAIL SURVEY

To address these questions, we emailed 10 recognized HIV vaccine experts, with the goal of eliciting a range for β from each of them. Of our 10 experts, only two have had extensive statistical training. Each expert was sent the email and attachment (shown below) on October 21, 2005. A second, reminder/re-invitation email was sent November 21, 2005 to our five non-responding experts. Finally, a third email was sent in January to our three remaining non-responders. The following is the email sent to our experts:

> Dear Dr. ,
>
> Peter Gilbert, Devan Mehrotra, and I have been working on statistical methods to estimate the causal effect of vaccination on post-infection outcomes in preventative HIV vaccine efficacy trials. We plan to use these methods in the analysis of Merck/HVTN's ongoing phase IIb intermediate-sized efficacy trial. We are soliciting input from you and 9 other HIV vaccine experts to help plan for the Merck/HVTN analysis – Devan is the lead statistician on the trial. We plan to use the information you give us to refine the analysis measuring the effect of vaccination on set-point viral load. In addition, we hope to write a paper in a statistics journal that discusses the feasibility of performing our new statistical methods. As part of this paper, we would like to publish some of the information we get from you and our other experts and discuss implications this information

could have on the analysis. Specifics will be kept anonymous.

Consider the Merck/HVTN trial. Given two participants assigned placebo who become infected during the course of the trial: who do you believe would be more likely infected if, contrary to fact, they were assigned vaccine?

____ the person with the higher set-point viral load

____the person with the lower set-point viral load

(Set-point viral load is defined here as the average of the log10 HIV RNA plasma levels at approximately two and three months after infection diagnosis.)

We would like you to translate your belief into a range of possible odds ratios. Consider two people infected in the placebo arm with set-point viral loads of 4.0 and 5.0 log10 copies/ml (approximately corresponding to the 25th and 75th percentiles of the MACS cohort). Suppose these two people had instead been assigned vaccine. Then the odds of infection in the vaccine arm for the individual with the set-point viral load of 5.0 is ____ times the odds of infection for the individual with the set-point viral load of 4.0. Please fill in this blank, giving us both a plausible lower and upper limit for this odds ratio.

____ lower limit for the odds ratio

____ upper limit for the odds ratio

We realize that the answers to these questions are impossible to know for certain; we are asking for your opinion. We have attached additional information that further describes the problem and may help you select a plausible range. Of course, if you have any questions or comments, please send one of us an email.

Thank you very much. We look forward to your email response. Sincerely,

Figure 1 and the following additional information were attached to the email:

The Purpose of this Odds Ratio Range

As you know, Merck/HVTN is currently conducting a "proof-of-concept" efficacy trial of an HIV vaccine designed to elicit a cell mediated immune response. One of the primary goals of this trial is to estimate the effect of vaccination on post-infection set-point viral load. A natural analysis would be to compare the set-point viral loads between infected participants in the vaccine and placebo arms. However, a comparison of this type could be misleading because it is restricted to subjects who are selected based on a post-randomization event (infection). Those who become infected in the placebo arm may have different characteristics than those infected in the vaccine arm, and if these differing characteristics are correlated with viral load, then a test comparing viral loads between infected individuals in both arms may be capturing this correlation, rather than the causal effect of vaccination on viral load. In essence, by analyzing the

subgroup of infected subjects, one may lose the benefit of randomization. We refer to this as possible selection bias.

Our methods account for this potential selection bias using a sensitivity analysis approach: We perform an analysis assuming a particular amount of selection bias and then repeat the analysis under different, scientifically plausible assumptions for the amount of selection bias. We describe the amount of selection bias by thinking about the likelihood that someone would be infected in the vaccine arm, given they got infected in the placebo arm and their set-point viral load. For example, consider two people in the placebo arm who became infected during the course of the trial. Suppose one of them (John) had a set-point viral load of 4 log10 copies/ml, whereas the other (Bill) had a set-point viral load of 5. Which one of these individuals do you believe would have been more likely to be infected if they had been randomized to vaccine?

We would like you to put a numerical range on this belief, translating your belief into a range of possible odds ratios. Recall that the odds of infection is defined as

$$\text{odds of infection } = \frac{\text{probability of infection}}{1 - \text{ probability of infection}}.$$

The odds ratio is simply the odds of infection for individuals with viral load y divided by the odds of infection for individuals with viral load $y-1$. An odds ratio of b would have the following interpretation: given infection

in the placebo arm, the odds of infection if randomized to vaccine is b times larger for individuals with a one-log higher viral load. Returning to our example of John and Bill. Suppose the odds ratio is 2. This would imply that the odds that Bill would have been infected if randomized to vaccine are twice the odds that John would have been infected if assigned vaccine. On the other hand, an odds ratio of 1/2 would imply that the odds of Bill being infected if randomized to vaccine are half the odds of John being infected. An odds ratio of one implies that the odds of infection are equal for John and Bill.

This odds ratio is the sensitivity parameter used to describe possible amounts of selection bias in our analyses. We are asking you to give us what you believe is a plausible range for this odds ratio sensitivity parameter. We realize that this odds ratio is impossible to know for certain; we are asking for your opinion.

Some Thoughts that May Help Select the Odds Ratio Range

Figure 1 demonstrates how different choices for the odds ratio sensitivity parameter have different implications on the amount of selection bias. Each histogram represents the distribution of the set-point viral loads for participants in the placebo arm who become infected over the course of the trial. The shaded region represents those participants infected in the placebo arm who would have also been infected if randomized to

vaccine, as implied by the specified odds ratio sensitivity parameter. For example, if the odds ratio sensitivity parameter equals 0, then all the participants infected in the placebo arm with the lowest viral loads would have been infected if assigned vaccine. In contrast, an odds ratio of 1 implies that there is no selection bias, i.e., the distribution of viral loads for all infected placebos is the same as for those who would have been infected if assigned vaccine. (These plots were created assuming that 70% of individuals infected in the placebo arm would have been infected if randomized to vaccine. Consideration of plausible values for the odds ratio sensitivity parameter should not depend on the level of vaccine efficacy to prevent infection.)

It may also be helpful to provide some examples of what might give rise to odds ratios above or below 1. An individual with a higher viral load would be more likely to be infected if assigned vaccine if, for example, people with relatively strong immune systems tend to have lower viral loads and if the vaccine is more likely to protect these individuals from infection. If this is thought to be the case, then the odds ratio would be greater than 1. On the other hand, the individual with a smaller viral load would be more likely to be infected if assigned vaccine if, for example, it is believed that the vaccine prevents infection with relatively strong/virulent viruses better than it prevents infection with weaker/a-virulent viruses. In this situation the odds ratio would be less than 1.

There may be some measured covariates that explain some of the potential selection bias. For example, viral load set-points for women tend to be lower than those for men. Suppose that in actuality the vaccine reduces the risk of infection in men but not in women, but that the vaccine has no effect on viral load for either gender. In this case, there will be fewer males among the vaccine infectees compared with the placebo infectees, and in a naïve between-group comparison of viral loads of infectees, this post-randomization gender disparity could (incorrectly) lead us to conclude that the vaccine is causing lower viral loads. The reverse could happen if the vaccine lowers the risk of infection in women but not in men, but has no effect on viral load. If one did not adjust for gender when performing the analysis, then the former case corresponds to an odds ratio sensitivity parameter < 1 while the latter corresponds to an odds ratio > 1. If gender were the only source of selection bias and one adjusted for gender in the analysis, then there would be no more selection bias and one could proceed using an odds ratio sensitivity parameter $= 1$. However, because we are concerned with possible selection bias arising from unmeasured covariates such as host genetics or immune factors, for this exercise we would like you to give us a range for the odds ratio sensitivity parameter in the situation where we do not adjust for any covariates.

Although we do not claim to have produced the perfect survey, quite a bit of thought went into the text of both the email and supplemental material. We needed to provide enough detail so that the experts could understand the problem: why there could be selection bias, our analysis approach, the definition of our sensitivity parameter, examples of the implications of certain OR values on selection bias, and reasons why one might believe certain OR values. All of this needed to be explained on a level that a non-statistician could understand. For example, Figure 1 was initially constructed using density curves, but later changed because we felt our experts would be more familiar with histograms. However, at the same time we needed to find a balance between providing sufficient detail and overwhelming experts (making it less likely that they would complete the survey), confusing them with too much information, and/or biasing their responses. Many of these issues are discussed in the prior elicitation literature (Chaloner 1996, Kadane and Wolfson 1998). A medical student with an emphasis in HIV (not one of our 10 experts) filled out a pilot survey and gave many helpful suggestions used in the final version.

4. OUR EXPERTS' RESPONSES

Eight of our 10 experts responded. Of the 8 responders, all but one said that the person with the higher placebo set-point viral load would be more likely infected if, contrary to fact, they were assigned vaccine. The other responder did not mark either box, indicating that they believed either answer was equally likely. Our responders' odds ratio ranges are given in Figure 2. Consistent with their answers to the first

question, most of the odds ratio ranges are greater than one.

Many of our experts' comments were quite informative and indicated that they understood the counterfactual nature of the problem. For example, one expert who gave an odds ratio range of 1-5 wrote:

> I believe there is first a "threshold" effect for becoming infected. Since most exposures do not result in persistent infection, there is presumably a selection process going on each time one is exposed to the virus. Viral variants more suited to transmission (eg CCR5-tropic, high replication capacity, etc) are more likely to be successful, and, in a non-vaccinated subject, host defenses consist primarily of innate responses (epithelial barrier, natural killer cells, etc). I presume that in most cases, adaptive immune responses (HIV specific antibodies or T cells) are absent or present at very low levels (due to prior unsucessful exposures) in the non-vaccinated subject. Once persistent infection is established, a post-infection phase occurs where the viral setpoint is determined by a complex interaction of viral and host characteristics. Looking at the viral loads in placebo recipients post-infection, really only tells you about the second, postinfection phase since, by definition, transmission has occurred. Post-infection, it is hard to know whether viral or host factors are responsible for the differences. You would almost have to roll back time and give Bill's virus to John and vice versa to be sure. Since I can't do that, I would postulate that, on

average, placebo recipients with higher viral loads have viruses that are "more fit" than placebo recipients with lower viral loads. Consequently, those more "fit" viruses would be more likely to be transmitted.

That being said, I think (hope) having received the vaccine will raise the threshold for infection, so even if exposed to a more "fit" virus, a vaccinated subject will be less likely to have persistent infection than a placebo recipient. This also implies that if a virus does slip by the adaptive immune responses elicited by the vaccine, it may be a more "fit" virus and therefore harder to control.

One responder gave us an odds ratio range of 1 to 3, but then added the following:

Although I gave answers, my feeling is that it would be people with VL in the lowest quartile (under placebo) that would likely have differential risk for infection if vaccinated... I actually don't think that those w/ VL of 4 vs 5 logs would be that different. I'd feel a bit better about my answers if they applied to people w/ VL of 2.5 and 4 logs (vs 4 and 5 logs).

This would suggest a different model than that given in A.4. After additional email discussion with this expert, we arrived at the following model:

$$P(S(1) = 1|S(0) = 1, Y(0) = y) = [1 + exp(-\alpha - \beta y^{I(y \leq 4)} 4^{I(y>4)})]^{-1}, \qquad (2)$$

where $exp(\beta)$, is between 1 and 3, and $I(\cdot)$ is the indicator function.

Only one expert responded in a manner indicative that he/she did not understand the odds ratio to be elicited. This expert suggested that we discuss the question by telephone, which we did, and by the end of our conversation we felt our expert understood the sensitivity parameter. From three of our experts we did not receive an essay response but only numerical values, so it is difficult for us to conclude whether or not these individuals understood the sensitivity parameter – we can only look at their ranges and compare them to the ranges of others.

Four of our experts commented on how mentally demanding it was to provide a plausible range for the counterfactual sensitivity parameter. After giving a thoughtful range and explanation, one expert sent us a second email titled "Friendly Revenge" with the first line reading: "After heavily concentrating [for a couple of hours] on your abstract questions I am getting back at you, who navigate in the abstract for a living, with the attached puzzle [a *Sudoku*]."

5. RANGES APPLIED TO A PRIOR HIV VACCINE TRIAL

The implications of these selected ranges on the Merck/HVTN analysis are not yet known, as the trial is still underway. However, for purpose of illustration we show a sensitivity analysis using these ranges applied to the non-white cohort of a previous HIV vaccine trial, VaxGen's trial of AIDSVAX B/B. Details and results from this trial are shown elsewhere (Flynn et al. 2005; Gilbert et al. 2005). In short, this was a double blinded trial from 1998-2003 that randomized over 5400 healthy, high-risk, HIV negative volunteers to vaccine or placebo in a 2:1 ratio. The vaccine

was found not to be effective at preventing against HIV infection, although subgroup analyses suggested that the vaccine may partially protect against infection for non-whites. (This latter conclusion has generally been discounted.) Of 914 non-whites in the trial, 30/604 (5.0%) in the vaccine arm became infected whereas 29/310 (9.4%) became infected in the placebo arm.

Figure 3 shows a sensitivity analysis of ACE, the mean difference in set-point viral loads in the non-white cohort of the VaxGen trial. The shaded regions of Figure 3 correspond to those ranges of the odds ratio sensitivity parameter that our experts believed most likely. Our elicited ranges are for the Merck/HVTN trial – a completely different vaccine – so applying these ranges to the VaxGen analysis is for illustrative purposes only. Notice that for odds ratios greater than 3, the 95% confidence interval for ACE does not include 0, suggesting that the vaccine caused lower viral loads. Five of our eight responding experts included odds ratios greater than 3 in their plausible range. However, three of these experts also included odds ratios less than 3 in their plausible range, where there is insufficient evidence to conclude that the vaccine is having an effect on viral loads. In summary, based on the ranges provided, 2 experts would have concluded that the vaccine was causing lower viral loads; 3 experts would have concluded that there was insufficient evidence to say the vaccine was impacting viral load; and 3 experts would have had conflicting conclusions within their chosen sensitivity parameter range.

The entire range of plausible values proposed by our respondents was $[0.7, 10]$.

A 95% confidence interval for ACE simultaneous for odds ratios over this range is $(-1.34, 0.43)$, corresponding to the 2.5th bootstrap percentile of $\widehat{ACE}(\beta)$ for $e^{\beta} = 10$ and the 97.5th bootstrap percentile of $\widehat{ACE}(\beta)$ for $e^{\beta} = 0.7$, respectively.

One could also imagine obtaining a single estimate for ACE by putting a distribution on the odds ratio sensitivity parameter based on the given ranges and then integrating over these values. Or one could do a fully Bayesian analysis defining a prior for the odds ratio sensitivity parameter using the experts' ranges. However, we prefer publishing the sensitivity analysis over the entire range of values, allowing other experts to see what would occur according to their own beliefs about the range for the odds ratio.

The analysis shown in Figure 3 includes the range given by our expert who proposed (2), a different model for $P(S(1) = 1|S(0) = 1, Y(0) = y)$. Figure 4 is a sensitivity analysis using this expert's model and odds ratio range [1, 3]. Similar to Figure 3, the null is not rejected in this range. Notice that the range for ACE is more narrow in Figure 4 than in Figure 3 over the same range of $exp(\beta)$. Under (2) the interpretation of β is technically the following: Given infection in the placebo arm, the odds of infection if randomized to the vaccine arm for $Y(0) = y_1$ versus $Y(0) = y_2$ are $exp\{\beta[min(y_1, 4) - min(y_2, 4)]\}$. Hence, β under (2) implies less selection bias than the same value of β under A.4.

6. DISCUSSION

Based on our informal survey, we believe elicitation of meaningful ranges for causal sensitivity parameters is feasible. We believe our subject-matter experts understood the sensitivity parameter and we find it interesting that our responders' ranges are broadly consistent. When analyzing data from the Merck/HVTN trial we will focus on estimates obtained by assuming odds ratios within the ranges given.

Critical readers may ask some of the following questions:

Did we bias our experts? Pascal wrote, "How difficult it is to submit anything to the judgement of another, without prejudicing his judgment by the manner in which we submit it!" (Pascal 1660) Although we provided examples both of situations producing odds ratios above and below one, perhaps some of our experts favored an odds ratio greater than one because they found the corresponding example more plausible. In general, it is difficult to find a balance between thoroughly explaining the sensitivity parameter and unknowingly providing our own biases. As pointed out in the prior elicitation literature, feedback and confirmation of ranges under different scenarios are important aspects of elicitation (Kadane and Wolfson 1998). For this report, however, we chose to sacrifice extensive feedback for the purpose of minimizing bias and to assess the feasibility of choosing a range for the sensitivity parameter with little coaching.

Did our experts really understand the odds ratio? Many have written about the challenges of interpreting odds ratios (Greenland 1987; Sackett, Deeks, and Altman

1996; Davies, Crombie, and Tavakoli 1998). Most medical researchers are at least generally familiar with odds ratios, and many could agree that an odds ratio of 1.1 is small, 3 is big, and 10 is very big. However, it is difficult to intuitively understand a specific odds ratio value. To avoid this problem, we considered transforming our odds ratio sensitivity parameter into a relative risk parameter (Scharfstein et al., 2006) or $P(S(1) = 1|S(0) = 1, Y(0) = y)$. However, to convert odds ratios to either of these quantities requires information about the underlying efficacy of the vaccine and the placebo viral load distribution. We decided that converting odds ratios to relative risks would add more complexity in interpretation and elicitation than simply using an odds ratio. In contrast, after a study has been completed, one can use estimated quantities to convert odds ratios to relative risks, if desired. Although we believe it is important to elicit sensitivity parameters a priori, we realize that the general acceptance or rejection of conclusions will be based on widespread agreement about a specific range, so it may be important to convert odds ratios to other quantities for better interpretation.

What if our subject matter experts are all wrong and we are misled? Kadane and Wolfson (1998) wrote: "Experts should be asked to assess only observable quantities." Cox (1998) added: "At one level direct elicitation of experts' *opinions* seems a bad idea likely to perpetuate the errors of the past. On the other hand elicitation of experts' *knowledge* and analytical processes is crucial. . . ." Counterfactuals are not observed; therefore the plausible ranges over which these analyses are performed are

indeed based on expert opinion. For this reason we suggest showing results under all possible values of the sensitivity parameter (in our case ranging the odds ratio from 0 to ∞). Then one can focus on results under certain values of the sensitivity parameter deemed more plausible (i.e., the elicited ranges).

In conclusion, we believe that elicitation of counterfactual sensitivity parameters is feasible. Sensitivity analyses that show answers over varying levels of assumptions are honest, informative, and can be interpreted by non-statisticians. We encourage their use.

ACKNOWLEDGEMENTS

We would like to thank our 8 responding HIV vaccine experts and VaxGen for allowing us to use their data. This work was supported in part by the Vanderbilt Vaccine Center and the Vanderbilt-Meharry Center for AIDS Research (NIH program P30 AI 54999).

REFERENCES

Chaloner, K. (1996), "The elicitation of prior distributions," In *Case Studies in Bayesian Biostatistcs* (eds D. Berry and D. Stangl), New York: Dekker, pp. 141–156.

Copas, J.B. and Li, H.G. (1997), "Inference for non-random samples," *Journal of the Royal Statistics Society, Series B*, 59, 55–95.

Cox, D.R. (1958), *Planning of Experiments*. New York: Wiley.

Cox, D.R. (1998), "Discussion on the papers on 'Elicitation'. *The Statisitician* **47,** 55–68.

Davies, H.T., Crombie, I.K., Tavakoli, M. (1998), "When can odds ratios mislead?" *British Medical Journal*, 316, 989–991.

Flynn, N.M., Forthal, D.N., Harro, C.D., et al. (2005), "Placebo-controlled trial of a recombinant glycoprotein 120 vaccine to prevent HIV infection," *Journal of Infectious Diseases*, 191, 645–665.

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

Gilbert, P.B., Ackers, M.L., Berman, P.W., et al. (2005), "HIV-1 virologic and immunologic progression and antiretroviral therapy initiation among HIV-1-infected participants in an efficacy trial of recombinant glycoprotein 120 vaccine," *Journal of Infectious Diseases*, 192, 974–983.

Gilbert, P.B., Bosch, R.J., and Hudgens, M.G. (2003), "Sensitivity analysis for the assessment of causal vaccine effects on viral load in HIV vaccine trials," *Biometrics*, 59, 531–541.

Greenland, S. (1987), "Interpretation and choice of effect measures in epidemiologic analyses," *American Journal of Epidemiology*, 125, 761–768.

Greenland, S. (1996), "Basic methods for sensitivity analysis of biases," *International Journal of Epidemiology*, 25, 1107–1116.

Greenland, S. (2005), "Multiple-bias modelling for analysis of observational data," *Journal of the Royal Statistical Society, Series A*, 168, 267–306.

Holland, P.W. (1986), "Statistics and causal inference," *Journal of the American Statistical Association*, 81, 945–960.

Kadane, J.B. and Wolfson, L.J. (1998), "Experiences in elicitation," *The Statistician*, 47, 3–19.

Leamer, E.E. (1974), "False models and post-data model construction," *Journal of the American Statistical Association*, 69, 122–131.

Mehrotra, D.V., Li, X., Gilbert P.B. (2006), "A comparison of eight methods for the dual-endpoint evaluation of efficacy in a proof-of-concept HIV vaccine trial," *Biometrics*, in press.

O'Hagan, A. (1998), "Eliciting expert beliefs in substantial practical applications," *The Statistician*, 47, 21–35.

Pascal, B. (1660), *Pensées*, Translated by A.J. Krailsheimer, London: Penguin Books, 1995, Number 105.

Robins, J.M. (1986), "A new approach to causal inference in mortality studies with

sustained exposure periods - Application to control of the healthy worker survivor effect," *Mathematical Modeling*, 7, 1393–1512.

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. (1987), "Sensitivity analysis for certain permutation inferences in matched observational studies," *Biometrika*, 74, 13–26.

Rosenbaum, P.R. (1988), "Sensitivity analysis for matching with multiple controls," *Biometrika*, 75, 577–581.

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.

Rubin, D.B. (2000), Comment on "Causal inference without counterfactuals," by A.P. Dawid. *Journal of the American Statistical Association* , 95, 435–437.

Sackett, D.L., Deeks, J.J., Altman, D.G. (1996), "Down with odds ratios!" *Evidence-Based Medicine*, 1, 164–166.

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

Scharfstein, D.O., Halloran, M.E., Chu H., Daniels, M.J. (2006), "On Estimation of Vaccine Efficacy Using Validation Samples with Selection Bias," *Biostatistics*, Epub ahead of print.

Shepherd, B.E., Gilbert, P.B., Jemiai, Y., and Rotnitzky, A. (2006), "Sensitivity analyses comparing outcomes only existing in a subset selected post-randomization, conditional on covariates, with application to HIV vaccine trials," *Biometrics*, 62, 332–342.

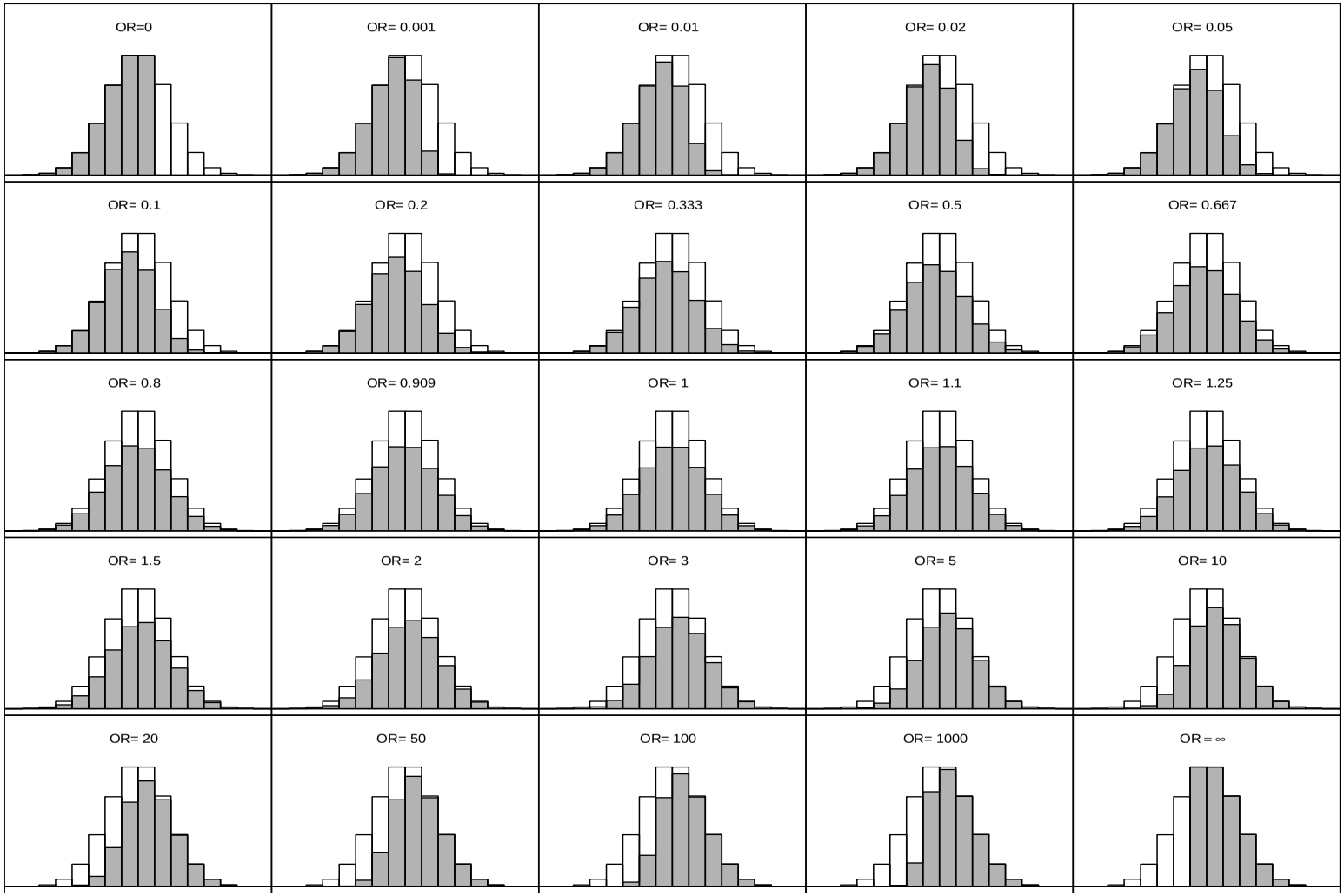

Figure 1: Each histogram represents the distribution of set-point viral loads for individuals randomized to placebo who became infected during the trial. The shaded portion of the histogram represents those individuals who would also have been infected if randomized to the vaccine arm, as determined by the odds ratio sensitivity parameter (OR). These plots show OR ranging from 0 (upper left plot) to ∞ (lower right plot), representing the selection bias extremes; OR=1 (middle plot) corresponds to no selection bias.

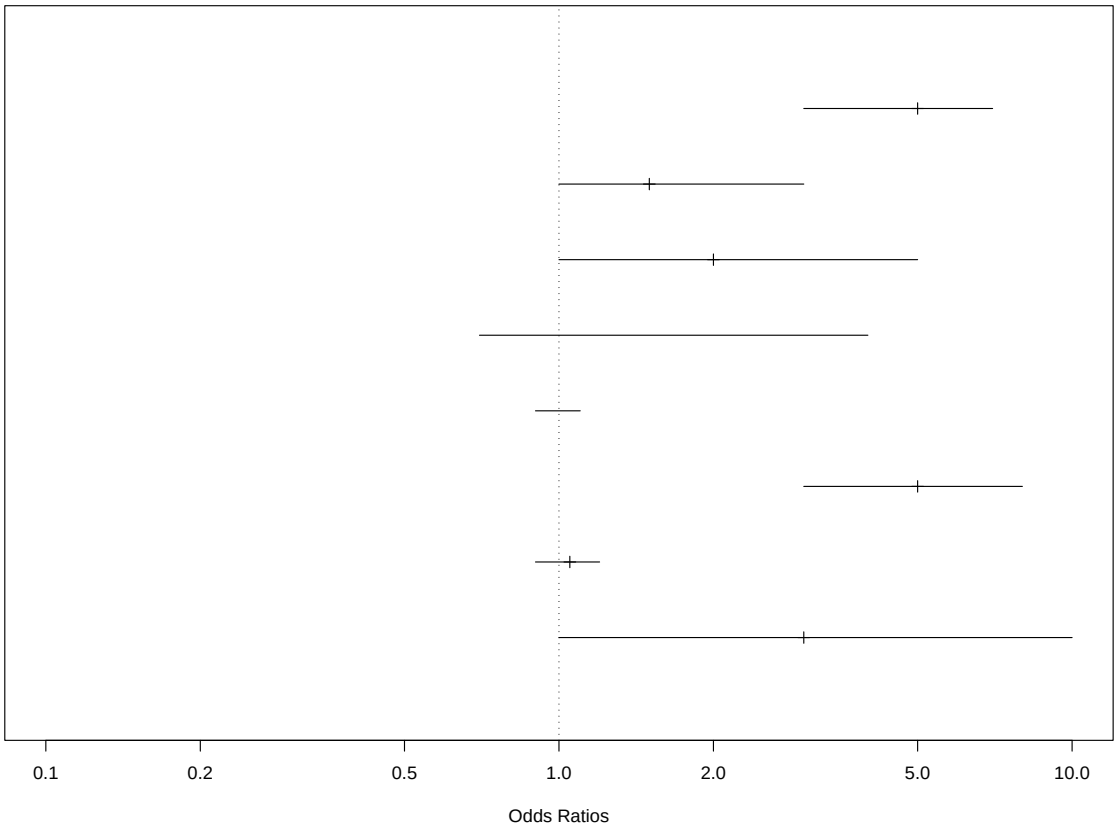

Figure 2: Ranges for the odds ratio sensitivity parameter, $\exp(\beta)$, elicited from our 8 responding HIV vaccine experts. Most experts also provided their most likely value, which is shown.

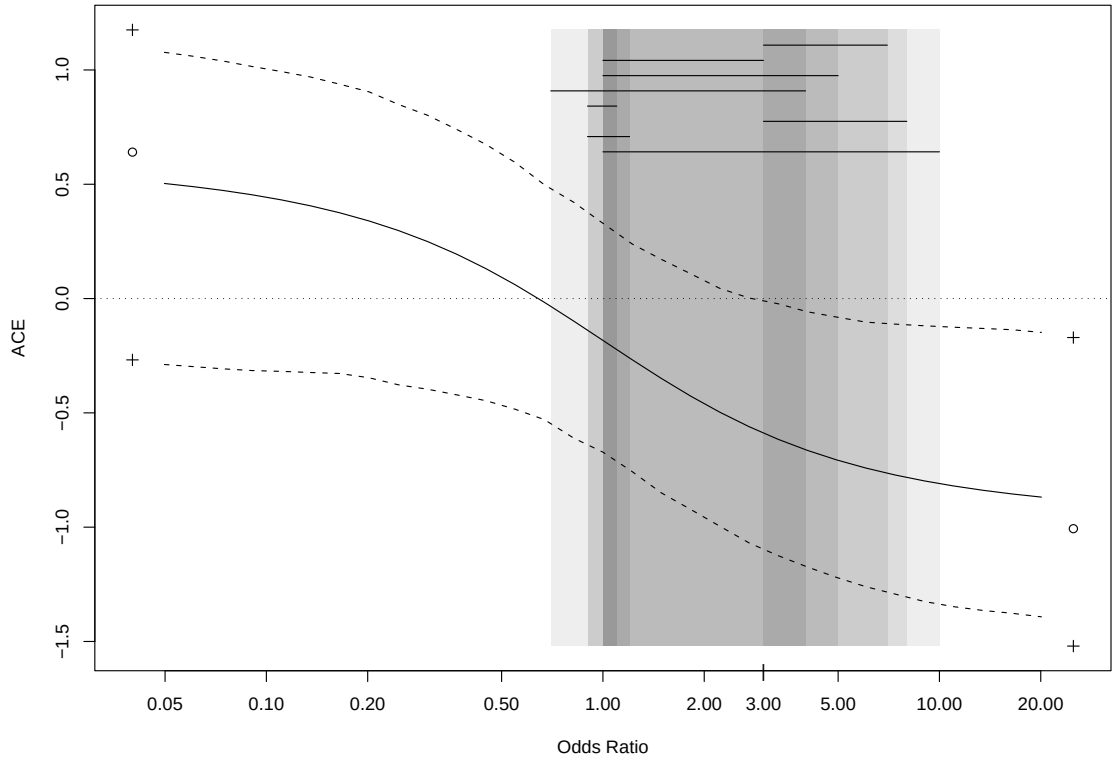

Figure 3: Sensitivity analysis of ACE in the non-white cohort of the VaxGen trial. The solid curve is the estimate of ACE as a function of the odds ratio sensitivity parameter, the dashed curves are the 95% pointwise confidence intervals based on the percentiles of 1000 bootstrap replications. The circles represent the estimates of ACE for selection bias odds ratios of 0 and ∞, and the crosses are their 95% confidence intervals. The shaded regions correspond to the ranges of plausible values for the selection bias odds ratio elicited from our experts for a different trial, the Merck/HVTN trial (individual ranges shown with horizontal lines); the darker shades signify values of the sensitivity parameter with greater overlap between experts.

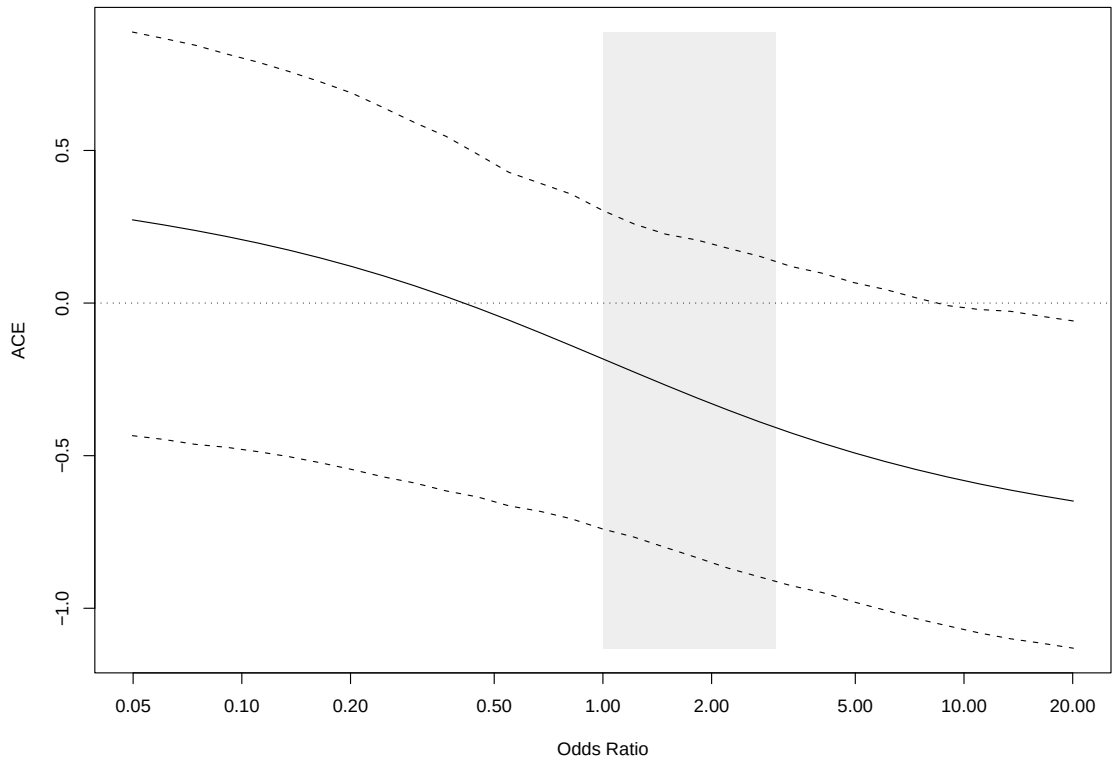

Figure 4: Sensitivity analysis of ACE in the non-white cohort of the VaxGen trial using model (2) proposed by one of our experts. The solid curve is the estimate of ACE, the dashed curves are the 95% pointwise confidence intervals based on the percentiles of 1000 bootstrap replications. The shaded region corresponds to our expert's range of plausible values.