Published in final edited form as:
Biostatistics. 2006 October ; 7(4): 615–629. doi:10.1093/biostatistics/kxj031.

On estimation of vaccine efficacy using validation samples with selection bias

DANIEL O. SCHARFSTEIN*,
Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205, USA

M. ELIZABETH HALLORAN,
Program in Biostatistics and Biomathematics, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA and Department of Biostatistics, University of Washington, Seattle, WA 98195, USA

HAITAO CHU, and
Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205, USA

MICHAEL J. DANIELS
Department of Statistics, University of Florida, Gainesville, FL 32611, USA

SUMMARY

Using validation sets for outcomes can greatly improve the estimation of vaccine efficacy (VE) in the field (Halloran and Longini, 2001; Halloran *and others*, 2003). Most statistical methods for using validation sets rely on the assumption that outcomes on those with no cultures are missing at random (MAR). However, often the validation sets will not be chosen at random. For example, confirmational cultures are often done on people with influenza-like illness as part of routine influenza surveillance. VE estimates based on such non-MAR validation sets could be biased. Here we propose frequentist and Bayesian approaches for estimating VE in the presence of validation bias. Our work builds on the ideas of Rotnitzky *and others* (1998, 2001), Scharfstein *and others* (1999, 2003), and Robins *and others* (2000). Our methods require expert opinion about the nature of the validation selection bias. In a re-analysis of an influenza vaccine study, we found, using the beliefs of a flu expert, that within any plausible range of selection bias the VE estimate based on the validation sets is much higher than the point estimate using just the non-specific case definition. Our approach is generally applicable to studies with missing binary outcomes with categorical covariates.

Keywords

Bayesian; Expert opinion; Identifiability; Influenza; Missing data; Selection model; Vaccine efficacy

1. Introduction

Many statistical methods have been developed to deal with estimating the causal effects of randomized treatments when outcomes are missing on some participants. Most methods rely on the non-identifiable assumption that the outcome of interest is missing at random (MAR)

*To whom correspondence should be addressed. dscharf@jhsph.edu.

Conflict of Interest: None declared.

(Little and Rubin, 2002). If the outcome is not MAR, then effect estimates could be subject to selection bias. Rotnitzky *and others* (1998, 2001), Scharfstein *and others* (1999), and Robins *and others* (2000) developed a frequentist selection model that displays the sensitivity analysis over a plausible range of selection bias parameters. Scharfstein *and others* (2003) developed a Bayesian approach that allows the formal incorporation of prior beliefs about the degree of the selection bias to obtain the full posterior distribution, a single summary of the sensitivity analysis. Others, for example Little (1994) and Daniels and Hogan (2000), have developed pattern-mixture models for sensitivity analyses. Some other approaches include the work of Baker *and others* (2003), Molenberghs *and others* (2001), Verbeke *and others* (2001), and Vansteelandt *and others* (2006).

A particular case of missing data occurs if the outcome of interest is difficult or expensive to ascertain, so that a surrogate outcome may be used instead. The outcome of interest may be measured on some of the study participants in a subset called a validation sample, while the surrogate is measured on all participants. In this situation, statistical missing data methods are available to use the outcomes of interest in the validation sample to correct the bias based on the non-specific case definition alone (Pepe *and others*, 1994). Halloran and Longini (2001), Halloran *and others* (2003), and Chu and Halloran (2004) have demonstrated the potential use of these methods for estimating vaccine efficacy (VE) on the example of an influenza vaccine. In a randomized study with a planned random sample selected for the validation set, MAR would be a reasonable assumption. However, in many situations, the selected sample may be a convenience sample, so that MAR is unlikely to hold. Halloran *and others* (2003) presented a simple model to explore the sensitivity of the VE estimates to the magnitude of the departure from the MAR assumption. However, their approach was ad hoc and did not give confidence bounds on their estimators.

Here, we formulate a class of selection models, indexed by interpretable parameters, to evaluate the sensitivity to selection bias when using validation sets to estimate VE. Frequentist and Bayesian approaches to inference will be presented. In developing and applying our methodology to the re-analysis of the influenza vaccine study, we worked closely with a scientific expert. Our approach is generally applicable to missing binary outcomes with categorical covariates.

2. Influenza vaccine study

A field study of a trivalent, cold-adapted, influenza virus vaccine (CAIV-T) was conducted in Temple-Belton, Texas, and surrounding areas during the 2000–2001 influenza season. The field study was part of a larger community-based, non-randomized, open-label field study conducted from 1998–2001 (Piedra *and others*, 2001; Gaglani *and others*, 2003). In Temple-Belton, eligible healthy children and adolescents aged 18 months through 18 years were offered CAIV-T vaccine through the Scott & White (S & W) Clinics from 1998–2001. The analysis includes children who were S & W Health Plan members, and is concerned with the CAIV-T vaccinations administered in the influenza season 2000–2001. Children received a single dose of CAIV-T each year that they enrolled. The primary clinical outcome was a non-specific case definition called medically attended acute respiratory infection (MAARI), which included all ICD-9-CM diagnoses codes (Codes 381–383, 460–487) for upper and lower respiratory tract infections, otitis media and sinusitis. Any individual presenting with history of fever and any respiratory illness at S & W Clinics was eligible to have a throat swab (or nasal wash in young infants) for influenza virus culture. The decision to obtain specimens was made irrespective of whether a patient had received CAIV-T. The specific case definition is culture-confirmed influenza. Table 1 contains the data. The overall fraction of MAARI cases sampled was a little higher in the unvaccinated than in the vaccinated groups ($p = 0.03$).

Halloran *and others* (2003) analyzed the data by adapting the mean score method for validation sets (Pepe *and others*, 1994) with the goal of evaluating the protective VE of CAIV-T vaccination in healthy children during the influenza season 2000–01. Chu and Halloran (2004) developed a Bayesian method. The overall vaccine effectiveness estimate based on the non-specific case definition was 0.18 (95% CI: 0.11, 0.24). The overall efficacy estimates incorporating the surveillance cultures using the mean score method was 0.79 (95% CI: 0.51, 0.91) and the Bayesian method was 0.74 (95% HPD: 0.50, 0.88). In this situation, using the surveillance cultures as a validation set resulted in a four-fold increase in estimates, much closer to the efficacy estimate of 0.93 (95% CI: 0.88, 0.97) obtained in a double-blind, randomized controlled trial (Belshe *and others*, 1998).

In the influenza vaccine study in Texas, no surveillance cultures were positive in the age group 1.5–4 years. In Halloran *and others* (2003), a continuity correction of 0.5 was added to the number of cultured samples and to the number positive in that age group in the mean score analysis. For this age group, their estimate of VE using the mean score method was 0.91 (95% CI: −0.24,0.99). The Bayesian method of Chu and Halloran (2004) yielded an estimate of 1.00 (95% HPD: 0.52,1.00). So, the Bayesian method provided a much tighter measure of uncertainty than the mean score method with the continuity correction.

The results of Halloran *and others* (2003) and Chu and Halloran (2004) are valid only if the culture-confirmed influenza status is MAR. In consulting with influenza experts, we learned that this assumption can easily be violated in this study if physicians tend to select children whom they believe to have influenza for culturing. Our goal is to develop frequentist and Bayesian methods for sensitivity analyses for these and similar data. Further, we develop a fully Bayesian procedure that formally incorporates expert beliefs about the culturing mechanism.

3. Notation and data structure

In the vaccine field study, let n be the total number of participants, and n_0 and n_1 the number of non-vaccinated and vaccinated participants, respectively. Let Z denote the vaccination indicator, taking on the value 1 if a participant is vaccinated and 0 if not vaccinated. Let $A(0)$ and $A(1)$ denote the indicator of MAARI (1: yes, 0: no) for a participant if she had been, possibly contrary to fact, unvaccinated or vaccinated, respectively. The observed MAARI outcome $A = A(Z)$ is observed for every participant. Let $Y(0)$ and $Y(1)$ denote influenza status (1: positive, 0: negative) for a participant if she had been, possibly contrary to fact, unvaccinated and vaccinated, respectively. Only one of these outcomes can be potentially observed. In this study, influenza status is biologically confirmed by a culture. In the validation substudy, a possibly non-random sample of the participants are biologically confirmed, so that influenza status, $Y = Y(Z)$, is known for a subset of the participants. Let R be the validation indicator, where $R = 1$ if sampled for validation and $R = 0$, otherwise. Sampling for validation only occurs for those with $A = 1$. Let X denote age category (0: 1.5–4 years, 1: 5–9 years, 2: 10–18 years) measured at the time of study entry.

With this notation, the observed data for an individual are $O = (Z, X, A, R, Y: A = R = 1)$. We assume that we observe n i.i.d. copies, $\mathbf{O} = \{O_i : i = 1, \ldots, n\}$.

Throughout, probabilities P, indexed by subgroup subscripts indicate restriction to the associated subpopulation. For example, for events A and B, $P_{z,x}[A] = P[A|Z = z, X = x]$ and $P_{z,x}[A|B] = P[A|B, Z = z, X = x]$.

4. Vaccine efficacy

The scientific goal is to use the observed data to estimate the causal effect of vaccination on the outcome Y, within age levels as well as overall. Specifically, we want to estimate age-specific VE

$$\mathrm{VE}_{S,x} = 1 - \frac{P_x[Y(1)=1]}{P_x[Y(0)=1]}$$

and overall VE

$$\mathrm{VE}_S = 1 - \frac{\Sigma_{x=0}^{2} P_x[Y(1)=1]\,P[X=x]}{\Sigma_{x=0}^{2} P_x[Y(0)=1]\,P[X=x]}.$$

To identify $\mathrm{VE}_{S,x}$, it is sufficient to identify $P_x[Y(z)=1]$ for $z = 0, 1$. For VE_S, we must identify $P_x[Y(z)=1]$ for all z and x, and the marginal distribution of X. While the marginal distribution is identified from the observed data without additional assumptions, the conditional probabilities $P_x[Y(z)=1]$ will require non-identifiable assumptions.

5. Two structural assumptions

Before proceeding further, we will make two structural assumptions to facilitate identification of $P_x[Y(z)=1]$.

Assumption 5.1

Z is independent of $\{A(0), A(1), Y(0), Y(1)\}$ given X.

This assumption states that vaccination status is independent of the potential outcomes $\{A(0), A(1), Y(0), Y(1)\}$, given age ($X$). That is, within levels of age, vaccination is randomized. Our expert felt that this assumption was reasonable, since there is no reason to believe that the decision to vaccinate was based on anything related to the potential outcomes.

Assumption 5.2

$A(z) = 0$ implies $Y(z) = 0$.

We make this assumption because the study design called for passive case ascertainment. As a result, the data structure is such that no participants who do not appear in the clinic are cultured for confirmation of influenza infection. The above assumption states that if a participant, under vaccination status z, does not have MAARI, then she does not have medically attended influenza. The interest is in efficacy against medically attended, culture-confirmed influenza, not influenza infection.

6. Identification of $P_x[Y(z)=1]$

With these assumptions, we can write

$$P_x[Y(z)=1] = P_{z,x}[Y(z)=1] \tag{6.1}$$

$$= P_{z,x}[Y=1] \tag{6.2}$$

$$= \sum_{r=0}^{1} P_{z,x}[Y=1|A=1,R=r]\, P_{z,x}[A=1,R=r]. \tag{6.3}$$

Equation (6.1) follows from randomization within levels of X, (6.2) uses the fact that $Y = Y(Z)$, and (6.3) follows from an application of the law of conditional probability and our second assumption above. Note that, for all z, x, r, $P_{z,x}[Y = 1|A = 1, R = 1]$ and $P_{z,x}[A = 1, R = r]$ are identifiable but $P_{z,x}[Y = 1|A = 1, R = 0]$ are not. Thus identification of $P_x[Y(z) = 1]$ will require identification of these latter probabilities.

The most common assumption employed to identify these probabilities is that of MAR (Little and Rubin, 2002). MAR states that R is independent of Y given (Z, A, X). This implies that, for all z, x, $P_{z,x}[Y = 1|A = 1, R = 0] = P_{z,x}[Y = 1|A = 1, R = 1]$. As a result, $P_x[Y(z) = 1]$ becomes identifiable. Since the assumption of MAR is untestable and is considered questionable by our scientific expert, it is useful to perform a sensitivity analysis to outcomes that are missing not at random.

7. Frequentist sensitivity analysis

7.1 Model specification

Scharfstein *and others* (1999, 2003) and Robins *and others* (2000) introduced a sensitivity analysis methodology in which a class of models (including MAR) are posited, each yielding identification of $P_{z,x}[Y = 1|A = 1, R = 0]$. They recommended that inferences about the estimands of interest be presented over a range of posited models, considered plausible by subject matter experts. In particular, they assumed a pattern-mixture model of the form

$$P_{z,x}[Y=1|A=1,R=0] = \frac{P_{z,x}[Y=1|A=1,R=1]\ \exp(\alpha_{z,x})}{c_{z,x}}, \tag{7.1}$$

where

$$c_{z,x} = P_{z,x}[Y=0|A=1,R=1] + P_{z,x}[Y=1|A=1,R=1]\ \exp(\alpha_{z,x})$$

and $\alpha_{z,x}$ is a specified non-identifiable constant to be varied in the sensitivity analysis. Setting $\alpha_{z,x} = 0$ is equivalent to MAR. Note that plugging the right-hand side of (7.1) into the right-hand side of (6.3) yields identification of $P_x[Y(z) = 1]$. Specifically,

$$P_x[Y(z)=1] = P_{z,x}[Y=1|A=1,R=1]\left\{P_{z,x}[A=1,R=1] + \frac{\exp(\alpha_{z,x})}{c_{z,x}} P_{z,x}[A=1,R=0]\right\}. \tag{7.2}$$

It is possible to re-write Model (7.1) as the following selection model:

$$\text{logit}\, P_{z,x}[R=0|A=1,Y=y] = h_{z,x} + \alpha_{z,x} y, \tag{7.3}$$

where

$$h_{z,x}=\log\left\{\frac{1}{c_{z,x}}\frac{P_{z,x}[R=0|A=1]}{P_{z,x}[R=1|A=1]}\right\}. \tag{7.4}$$

For subjects with $Z = z$, $X = x$, and MAARI, $\alpha_{z,x}$ is interpreted as the log odds ratio of being unvalidated for diseased vs. undiseased subjects. So, $\alpha_{z,x}$ positive or negative indicates that diseased subjects have lower or higher odds of being validated, respectively.

When eliciting plausible ranges for $\alpha_{z,x}$, our expert found it easier to think about selection bias on a relative risk as opposed to an odds ratio scale. Specifically, he felt more comfortable expressing opinions about the relative risk of being validated given that a MAARI participant has influenza, compared with having another influenza-like illness. As a result, we re-formulated the above models in terms of the relative risk selection bias parameters

$$\beta_{z,x}=\frac{P_{z,x}[R=1|A=1,Y=1]}{P_{z,x}[R=1|A=1,Y=0]}. \tag{7.5}$$

We can re-parametrize Model (7.3) in terms of $\beta_{z,x}$. Letting $\eta_{z,x} = P_{z,x}[R = 1|A = 1, Y = 0]$, we can show that there is a one-to-one mapping between $(h_{z,x}, \alpha_{z,x})$ and $(\eta_{z,x}, \beta_{z,x})$. Specifically,

$$\begin{aligned}\eta_{z,x} &= (1+\exp(h_{z,x}))^{-1}\\ \beta_{z,x} &= \frac{1+\exp(h_{z,x})}{1+\exp(h_{z,x}+\alpha_{z,x})}\end{aligned}$$

or

$$\begin{aligned}h_{z,x} &= \log(1-\eta_{z,x}) - \log(\eta_{z,x})\\ \alpha_{z,x} &= \log\left(\beta_{z,x}^{-1}-\eta_{z,x}\right) - \log(1-\eta_{z,x}).\end{aligned}$$

Because of this one-to-one relationship, specification of $\beta_{z,x}$ will lead to identification of $P_x[Y(z) = 1]$ via the following formula:

$$\begin{aligned}P_x[Y(z)=1] &= \frac{P_{z,x}[Y=1|A=1,R=1]P_{z,x}[A=1]}{\beta_{z,x}P_{z,x}[Y=0|A=1,R=1]+P_{z,x}[Y=1|A=1,R=1]}\\ &= \frac{P[Z=z,X=x,A=1,R=1,Y=1]P[Z=z,X=x,A=1]/P[Z=z,X=x]}{\beta_{z,x}P[Z=z,X=x,A=1,R=1,Y=0]+P[Z=z,X=x,A=1,R=1,Y1]}.\end{aligned} \tag{7.6}$$

7.2 Estimation and inference

The frequentist non-parametric estimator of $P_x[Y(z) = 1]$ can be found by replacing the probabilities P in (7.6) by their empiricals $\tilde{P}$. Then

$$\widehat{P}_x[Y(z)=1] = \frac{\tilde{P}[Z=z,X=x,A=1,R=1,Y=1]\,\tilde{P}[Z=z,X=x,A=1]/\tilde{P}[Z=z,X=x]}{\beta_{z,x}\tilde{P}[Z=z,X=x,A=1,R=1,Y=0]+\tilde{P}[Z=z,X=x,A=1,R=1,Y=1]}. \tag{7.7}$$

The right-hand side of the above equation reduces to the results with the mean score method when $\beta_{z,x} = 1$, for all z and x.

We can then estimate $VE_{S,x}$ by

$$\widehat{\mathrm{VE}}_{S,x}=1-\frac{\widehat{P}_x[Y(1)=1]}{\widehat{P}_x[Y(0)=1]},$$

and VE_S by

$$\widehat{\mathrm{VE}}_S=1-\frac{\Sigma_{x=0}^{2}\widehat{P}_x[Y(1)=1]\,\tilde{P}[X=x]}{\Sigma_{x=0}^{2}\widehat{P}_x[Y(0)=1]\,\tilde{P}[X=x]}.$$

In Section A of supplementary material available at *Biostatistics* online (http://www.biostatistics.oxfordjournals.org), we derive the large sample-based confidence intervals for $\mathrm{VE}_{S,x}$ and VE_S. The sensitivity analysis proceeds by varying the $\beta_{z,x}$ over plausible ranges. When a stratum has a relatively small number of validation cultures and small number of positive cultures, as in our influenza example, these large sample confidence intervals may not perform well. When there are no positive cultures, one can use the continuity correction approach of Halloran *and others* (2003). In Section 8, we show how to compute Bayesian credible intervals which should perform better in this setting.

8. Bayesian Inference

Scharfstein *and others* (2003) developed a Bayesian methodology that allows full posterior inference about the estimands of interest, by assuming informative prior distributions on the selection bias parameters on the odds ratio scale. Their work did not allow for covariates and assumed independent priors across treatment groups. Our goal is to extend their work to the relative risk parametrization of selection bias, discrete covariates, and dependence of the priors for the relative risk parameters across treatment groups.

To simplify notation, we let $\beta_z=(\beta_{z,0},\beta_{z,1},\beta_{z,2})'$, $\beta=\left(\beta_0',\beta_1'\right)$, $\eta_z=(\eta_{z,0},\eta_{z,1},\eta_{z,2})'$, $\eta=\left(\eta_0',\eta_1'\right)'$, $p_{z,x}=P_{z,x}[Y=1|A=1]$, $\boldsymbol{p}_z=(p_{z,0},p_{z,1},p_{z,2})'$, $\boldsymbol{p}=\left(\boldsymbol{p}_0',\boldsymbol{p}_1'\right)$, $\phi_{z,x,a}=P_z$, and $\phi=\left(\phi_0',\phi_1'\right)'$. With this notation,

$$P_x[Y(z)=1]=P_{z,x}[Y=1]=p_{z,x}\frac{\phi_{z,x,1}}{\Sigma_{a=0}^{1}\phi_{z,x,a}},$$

and

$$\mathrm{VE}_{S,x}=1-\frac{p_{1,x}}{p_{0,x}}\frac{\phi_{1,x,1}}{\phi_{0,x,1}}\frac{\Sigma_{a=0}^{1}\phi_{0,x,a}}{\Sigma_{a=0}^{1}\phi_{1,x,a}}, \tag{8.1}$$

$$\mathrm{VE}_S=1-\frac{\Sigma_{x=0}^{2}\left\{p_{1,x}\phi_{1,x,1}\left(\Sigma_{z=0}^{1}\Sigma_{a=0}^{1}\phi_{z,x,a}\right)/\left(\Sigma_{a=0}^{1}\phi_{1,x,a}\right)\right\}}{\Sigma_{x=0}^{2}\left\{p_{0,x}\phi_{0,x,1}\left(\Sigma_{z=0}^{1}\Sigma_{a=0}^{1}\phi_{z,x,a}\right)/\left(\Sigma_{a=0}^{1}\phi_{0,x,a}\right)\right\}}. \tag{8.2}$$

In the prior specification for $\boldsymbol{\beta}$, we provide two options: (1) Bayesian analogue of the frequentist sensitivity analysis and (2) fully Bayesian analysis. For option (1), we specify point-mass priors

on $\boldsymbol{\beta}$ and estimate the posterior distributions of the estimands of interest, over a range of $\boldsymbol{\beta}$. This approach is comparable to the frequentist sensitivity analysis described above, but does not rely on large sample approximations. For option (2), we specify a non-degenerate prior distribution on $\boldsymbol{\beta}$, elicited from a subject matter expert. This approach has the advantage of providing a single summary of the posterior inference about the estimands, which naturally incorporates the uncertainty due to selection bias.

8.1 Prior specification for (β′,η′, p′, φ′)′

In our specification of the joint prior distribution on $(\boldsymbol{\beta}', \boldsymbol{\eta}', \boldsymbol{p}', \boldsymbol{\varphi}')'$, we make the following assumptions:

- $\boldsymbol{\varphi}$ is independent of $(\boldsymbol{\beta}', \boldsymbol{\eta}', \boldsymbol{p}')'$.
- $\boldsymbol{\varphi}_1$ is independent of $\boldsymbol{\varphi}_0$.
- $\boldsymbol{p}$ is independent of $(\boldsymbol{\beta}', \boldsymbol{\eta}')'$.
- The components of $\boldsymbol{p}$ are independent.
- The components of $\boldsymbol{\beta}$ may be correlated.
- Given $\beta_{z,x}$, $\eta_{z,x}$ is independent of $\{\beta_{1-z,x'}, \eta_{1-z,x'} : x' \neq x\}$.
- • $\beta_{z,x}$ and $\eta_{z,x}$ are only related through $\beta_{z,x}$'s restriction on the support of $\eta_{z,x}$.

Thus, the prior distribution of $(\boldsymbol{\beta}', \boldsymbol{\eta}', \boldsymbol{p}', \boldsymbol{\varphi}')'$ can be written as

$$\pi(\beta, \eta, \boldsymbol{p}, \phi) = \pi(\beta) \prod_{z=0}^{1} \prod_{x=0}^{2} \pi(p_{z,x}) \, \pi(\eta_{z,x} | \beta_{z,x}) \, \pi(\phi_z).$$

We further assume that

- $\pi(\boldsymbol{\beta})$ is an informative prior on a compact subset of $\mathbb{R}^6$. For the fully Bayesian analysis of the vaccine trial, we assume an informative multivariate normal prior on the log $\boldsymbol{\beta}$ scale. The details of the informative priors are described in Section 9.1.
- $\pi(p_{z,x})$ is a uniform prior on [0, 1].
- $\pi(\eta_{z,x}|\beta_{z,x})$ is a uniform prior on $\left[0, \min\left(\beta_{z,x}^{-1}, 1\right)\right]$. Specifically,

  $$\pi(\eta_{z,x} | \beta_{z,x}) = \max\{\beta_{z,x}, 1\} \, I(\eta_{z,x} < \min\{1, 1/\beta_{z,x}\}),$$

  where $I(A)$ denotes the indicator of the event A.
- $\pi(\varphi_z)$ is a non-informative prior on the set

$$\left\{(\phi_{z,0,0}, \phi_{z,0,1}, \phi_{z,1,0}, \phi_{z,1,1}, \phi_{z,2,0}, \phi_{z,2,1}) : 0 \leq \phi_{z,x,a} \leq 1, \sum_{x=0}^{2} \sum_{a=0}^{1} \phi_{z,x,a} = 1\right\}.$$

  Assume that the combinations of (x, a) have $k = 1, \ldots, K$ categories with n_{zk} subjects in each category. The prior on $\boldsymbol{\varphi}_z$ is specified as a Dirichlet, $D(1, \ldots, 1)$. This represents a prior sample of K, the number of categories of (x, a). We could make this less informative by replacing the 1's with $1/K$.

8.2 Sampling from the posterior distribution

To sample from the posterior, we constructed a Gibbs sampling algorithm with data augmentation (Tanner and Wong, 1987) and slice sampling (Damien *and others*, 1999; Neal, 2003). Section B of the supplementary material available at *Biostatistics* online (http://www.biostatistics.oxfordjournals.org) provides a full description of the algorithm.

9. Data Analysis

9.1 Informative priors

For Bayesian inference, we specify informative priors for the selection bias relative risk parameters, $\boldsymbol{\beta}$, by age group and vaccination status. We asked an influenza expert the following question: "If a physician were doing surveillance cultures during an influenza season, what is the probability that he would select the children who actually had true influenza over the children who just had non-specific respiratory symptoms to culture?" He responded that this was very hard to answer. One "would be more likely to be correct in the unvaccinated," because unvaccinated children presenting with true influenza would have more typical, severe disease than the vaccinated children. One would be "less likely to be correct in young children under 5 years," because children under 5 years experience many other severe respiratory diseases that could be mistaken for influenza, while older children are already immune to such diseases. He added that the degree of selection bias would also "depend on the rules for collection, for example, a certain number per week or with specific symptoms."

Another influenza expert had similar views. He provided us with his best guess for each of the univariate relative risk selection bias parameters $\beta_{z,x}$ defined in (7.5) and his belief about the interval that would likely contain 90% of the prior distribution for each $\beta_{z,x}$. He also provided us with the degree of correlation of the selection bias by age group and vaccine status. Correlations could also be elicited graphically (Shardell, 2005). Although $\pi(\boldsymbol{\beta})$ is an informative prior on a compact subset of $\mathbb{R}^6$, the expert did not have different prior beliefs about the selection bias in the 5- to 9-year and 10- to 18-year age groups. Thus, we present the prior distributions for these two age groups as one group in Table 2.

For our analysis, we plugged the elicitations into a multivariate Normal prior on the log $\boldsymbol{\beta}$ scale as follows. We transformed the elicited best guesses for each $\beta_{z,x}$ and 90% range to the log $\beta_{z,x}$ scale. Using qnorm in Splus, we found the univariate Normal (mean, standard deviation) distributions that best approximated the elicited priors on the log scale. In the unvaccinated 5- to 18-year olds, the elicitation was quite consistent with a Normal distribution. In the other three groups, adjustments were necessary as shown in Table 2. The expert felt comfortable with the changes in elicited and proposed priors in light of the uncertainty about the selection bias. Table 2 contains the log-transformed elicited priors, the slightly adjusted univariate Normal priors with their 90% interval on the log $\beta_{z,x}$ scale, and the corresponding best guess and ranges on the relative risk, $\beta_{z,x}$, scale. The corresponding univariate distributions that we used for the log $\beta_{z,x}$ for the unvaccinated 1.5- to 4-year old and older children are Normal (0.69, 0.43) and Normal (1.10, 0.25) and for the vaccinated 1.5- to 4-year olds and older children are Normal (0.18,0.57) and Normal (0.53, 0.32).

The expert believed that the correlation in selection bias among the strata would be high. He suggested a correlation as high as 0.90. The corresponding covariance matrices for $\pi(\boldsymbol{\beta})$ were constructed from the marginal univariate Normal distributions and the correlations. We also performed the analysis assuming zero correlation.

All programs were written in R. The Markov Chain Monte Carlo algorithm was run for 500 000 iterations with 100 000 discarded as burn-in.

9.2 Results

Figures 1 and 2 display results of the frequentist sensitivity analysis. These figures present results by age stratum. The selection bias parameters were varied over the 90% ranges elicited from the expert. It was not feasible to present parsimoniously the overall results because the selection bias parameters were too high-dimensional. This is a key drawback of the frequentist sensitivity analysis methodology.

Figure 1 shows the estimated probabilities (and 95% confidence intervals) for influenza in the vaccinated and unvaccinated groups, for each age stratum, as a function of $\beta_{z,x}$. Figure 2 shows the point estimates and lower 95% confidence bounds for the age-group-specific efficacy. The black diamonds indicate the results at the best guess of the expert. Within these ranges and within each age group, the VE estimates based on the validation sets are higher than the point estimates based on the non-specific definition, which were 0.2, 0.25, and 0.14 for the age groups 1.5–4, 5–9, and 10–18, respectively. The lower confidence bounds indicate the degree of variability.

The prior and posterior distributions for the relative risks of selection $\beta_{z,x}$'s by vaccination status and age group are the same, since there is no information in the data about the degree of selection bias (figure not shown). Figure 3 shows the Bayesian posterior distribution of the age-group-specific efficacy and overall efficacy using the informative prior distributions from Table 2, assuming a correlation of 0.9. The mode is 1.00 in the age group 1.5–4 years, since there are no positive cultures in the vaccinated group in that age group. The results assuming a zero correlation were nearly identical (not shown).

Table 3 compares the summaries of the Bayesian posterior distributions and of the frequentist estimates and 95% confidence intervals. The assumption of MAR results in an overestimate of the VE compared with the selection bias relative risk assumptions elicited from the expert. Consistent with Chu and Halloran (2004), the Bayesian posterior means are somewhat lower than the frequentist estimates. In general, the Bayesian credible intervals are tighter than the frequentist confidence intervals. This is especially true for the age group 1.5–4 years, where the validation sample is small and there are no positive cultures.

It is interesting to note that the Bayesian analysis using an informative prior for $\boldsymbol{\beta}$ had a narrower credible interval than the analysis with $\boldsymbol{\beta}$ fixed at its best guess, i.e. a degenerate prior (see the first two rows of Table 3). At first glance, this former interval would be expected to be wider. However, the posterior results are a weighted average of posteriors with fixed $\boldsymbol{\beta}$, of which some have larger and some have smaller variation than the posterior at the best guess. Thus, it is plausible to get a smaller width for the overall posterior credible interval than that fixed at the best guess. Similar phenomena have been previously documented in Gustafson and Greenland (2006) in the context of adjusting for misclassification of exposure.

10. Discussion

In this paper, we have developed both frequentist and Bayesian methods for analyzing missing binary outcomes that are thought to be informatively (i.e. related to the unobserved outcome) missing. Our relative risk parametrization of selection bias, incorporation for discrete covariates, and prior dependence of the selection bias across treatment groups extends earlier work (Scharfstein *and others*, 2003), which used the log odds ratio parametrization without covariates. We elicited informative age group and vaccination status prior distributions from an influenza expert. Bayesian inference with informative priors on the selection bias parameters provides a useful and parsimonious way of drawing inference about VE that incorporates expert uncertainty about the missing data mechanism. In addition, the Bayesian approach provides better small-sample inference.

The frequentist or Bayesian sensitivity analysis approach provides much greater detail than the single summary from the fully Bayesian analysis. As a consequence, when the dimension of the selection parameters is greater than 3 or 4, it is harder to visualize. While the advantage of the Bayesian sensitivity analysis is the finite sample performance, it is computationally very intensive. The frequentist sensitivity analysis is computationally more feasible and will perform well when the sample sizes are large.

We used our proposed methods to re-analyze an influenza vaccine field study. We did a formal sensitivity analysis to evaluate the effects of preferential selection of children with non-specific illness for obtaining surveillance cultures to confirm true influenza. Our analysis showed that under plausible ranges of selection bias, the VE estimates, though lower than when assuming MAR, are substantially higher than those based on the non-specific influenza-like illness definition alone. Our methods will be generally useful in future vaccine field studies, or other similar studies, in which confirmatory biological specimens are not MAR.

For our development in this paper, we made the assumption that any person who does not present with non-specific influenza-like illness also does not have medically attended influenza illness. Implicitly, we assumed a degenerate prior at zero for the $P_{z,x}[Y = 1|A = 0]$. In many vaccine field studies, cases are ascertained on symptoms, then possibly confirmed biologically. Those participants without symptoms are generally not confirmed biologically. In the study analyzed here, ascertainment was passive through clinic visits, so our efficacy measure was for medically attended influenza illness. Ascertainment on symptoms could also be active, say, through regular phone calls to the home. Our methods could be used for studies with an active ascertainment method without further extension, whereby $A(z) = 1$ would denote being symptomatic by whatever case definition was used for that study, and the interpretation of VE would be for symptomatic influenza.

It is straightforward to extend our method to the situation that infection is confirmed, perhaps serologically, in a sample of people who did not have non-specific illness, $A(z) = 0$. In this case, we would not need Assumption 2. The scientific question then would be to estimate efficacy against infection, not medically attended disease, as in this paper. There would be additional selection bias parameters in that situation. However, if a study were well enough planned to sample asymptomatic participants, it would be hoped that the sampling would be planned to be random, so that selection bias would not play a role.

Our informative priors for the selection bias parameters were assumed to be multivariate Normal on the log $\boldsymbol{\beta}$ scale. Prior distributions could also be constructed less parametrically using the information elicited from the experts. When using our methods to analyze future studies, the providers making the decision from whom to obtain biological specimens could be directly asked their prior beliefs about selection bias.

In ongoing research, we are extending the Bayesian methods to incorporate higher-dimensional covariates. We are also working on methods for longitudinal and time-to-event outcomes.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Daniel Scharfstein was partially supported by National Institutes of Health (NIH) grants R01-CA85295, P30-MH066247, and R01-A132475. Elizabeth Halloran was partially supported by NIH grant R01-AI32042. Michael Daniels was partially supported by NIH R01-CA85295 and R01-HL079457. Haitao Chu was supported in part by the NIH through the data coordinating centers for the Multicenter AIDS Cohort Study (UO1-AI-35043), the Women's Interagency HIV Study (UO1-AI-42590), and the Pediatric Chronic Kidney Disease cohort study (UO1-DK-066116).

We thank Ira M. Longini, Jr., Pedro A. Piedra, W. Paul Glezen, and Arnold Monto for helpful comments during elicitation of the informative priors.

References

Baker SG, Ko C-W, Graudbard BI. A sensitivity analysis for nonrandomly missing categorical data arising from a national health disability survey. Biostatistics 2003;4:41–56. [PubMed: 12925329]

Chu H, Halloran ME. Estimating vaccine efficacy using auxiliary outcome data and a small validation set. Statistics in Medicine 2004;23:2697–712. [PubMed: 15316953]

Damien P, Wakefield J, Walker S. Gibbs sampling for Bayesian non-conjugate and hierarchical models by using auxiliary variables. Journal of the Royal Statistical Society, Series B 1999;61:331–44.

Daniels MJ, Hogan J. Reparametrizing the pattern mixture model for sensitivity analyses under informative dropout. Biometrics 2000;56:1241–8. [PubMed: 11129486]

Gaglani MJ, Piedra PA, Herschler GB, Griffith ME, Kozinetz CA, Riggs MW, Fewlass C, Halloran ME, Longini IM, Glezen P. Direct effectiveness of the intranaal, live-attenuated trivalent, cold-adapted, influenza virus vaccine (CAIV-T) against the 2000–2001 influenza A (H1N1) and B epidemic in healthy children. Archives of Pediatric and Adolescent Medicine 2003;158:65–73.

Gustafson P, Greenland S. Curious phenomena in Bayesian adjustment for exposure misclassification. Statistics in Medicine 2006;25:87–103. [PubMed: 16220473]

Halloran ME, Longini IM. Using validation sets for outcomes and exposure to infection in vaccine field studies. American Journal of Epidemiology 2001;154:391–8. [PubMed: 11532779]

Halloran ME, Longini IM, Gaglani MJ, Piedra PA, Chu H, Herschler GB, Glezen WP. Estimating efficacy of trivalent, cold-adapted, influenza virus vaccine (CAIV-T) against influenza A (H1N1) and B using surveillance cultures. American Journal of Epidemiology 2003;158:305–11. [PubMed: 12915495]

Little RJA. A class of pattern-mixture models for normal missing data. Biometrika 1994;81:471–83.

Little, RJA.; Rubin, DB. Statistical Analysis with Missing Data. Vol. 2nd edition. John Wiley and Sons; Hoboken, NJ: 2002.

Molenberghs G, Kenward MG, Goetghbeur EJT. Sensitivity analysis for incomplete contingency tables. Applied Statistics 2001;50:15–29.

Neal RM. Slice sampling (with discussion). Annals of Statistics 2003;31:705–67.

Pepe MS, Reilly M, Fleming TR. Auxiliary outcome data and the mean score method. Journal of Statistical Planning and Inference 1994;42:137–60.

Piedra, PA.; Gaglani, M.; Herschler, G.; Riggs, M.; Griffith, M.; Kozinetz, C.; Glezen, WP. Safety and effectiveness of trivalent, cold-adapted influenza vaccine (CAIV-T) in children. In: Osterhaus, ADME.; Cox, N.; Hampson, AW., editors. Options for Control of Influenza IV. Excerpta Medica; Amsterdam: 2001. p. 939-43.

Robins, JM.; Rotnitzky, A.; Scharfstein, DO. Sensitivity analysis for selection bias and unmeasured confounding in missing data and causal inference models. In: Halloran, ME.; Berry, DA., editors. Statistics in Epidemiology, Environment and Clinical Trials. Springer; New York: 2000. p. 1-94.

Rotnitzky A, Robins JM, Scharfstein D. Semiparametric regression for repeated outcomes with non-ignorable non-response. Journal of the American Statistical Association 1998;93:1321–39.

Rotnitzky A, Scharfstein D, Su T, Robins JM. Methods for conducting sensitivity analysis of trials with potentially non-ignorable competing causes of censoring. Biometrics 2001;57:103–13. [PubMed: 11252584]

Scharfstein DO, Daniels MJ, Robins JM. Incorporating prior beliefs about selection bias into the analysis of randomized trials with missing outcomes. Biostatistics 2003;4:495–512. [PubMed: 14557107]

Scharfstein DO, Rotnitzky A, Robins JM. Adjusting for nonignorable drop-out using semiparametric nonresponse models (with discussion). Journal of the American Statistical Association 1999;94:1096–146.

Shardell, M. The analysis of informatively coarsened discrete time-to-event data. Johns Hopkins University; Baltimore, MD: 2005. PhD. Thesis

Tanner MA, Wong WH. The calculation of the posterior distributions by data augmentation. Journal of the American Statistical Association 1987;82:528–40.

Vansteelandt S, Goetghebeur E, Kenward MG, Molenberghs G. Ignorance and uncertainty regions as inferential tools in a sensitivity analysis. Statistica Sinica. 2006in press

Verbeke G, Molenberghs G, Thijs H, Lesaffre E, Kenward MG. Sensitivity analysis for nonrandom dropout: a local influence approach. Biometrics 2001;57:7–14. [PubMed: 11252620]

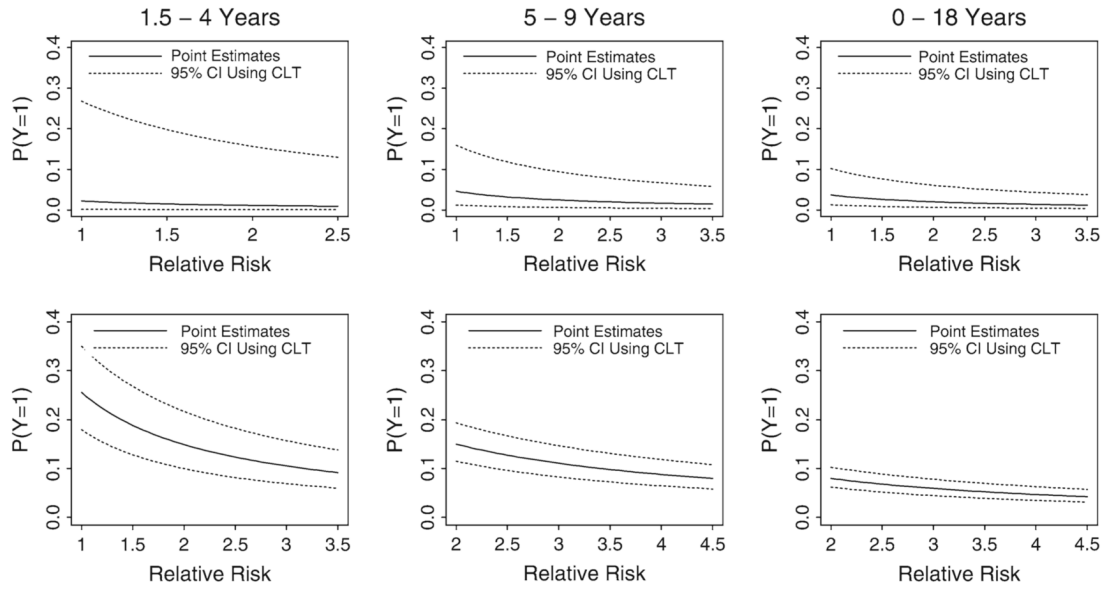

Fig. 1.

Frequentist sensitivity analysis of $P_{z,x}[Y = 1]$. Shown are the estimated probabilities (and 95% confidence intervals) for influenza in the vaccinated and unvaccinated groups, for each age stratum, as a function of the relative risk selection bias parameter $\beta_{z,x}$, varied over the 90% ranges elicited from the expert.

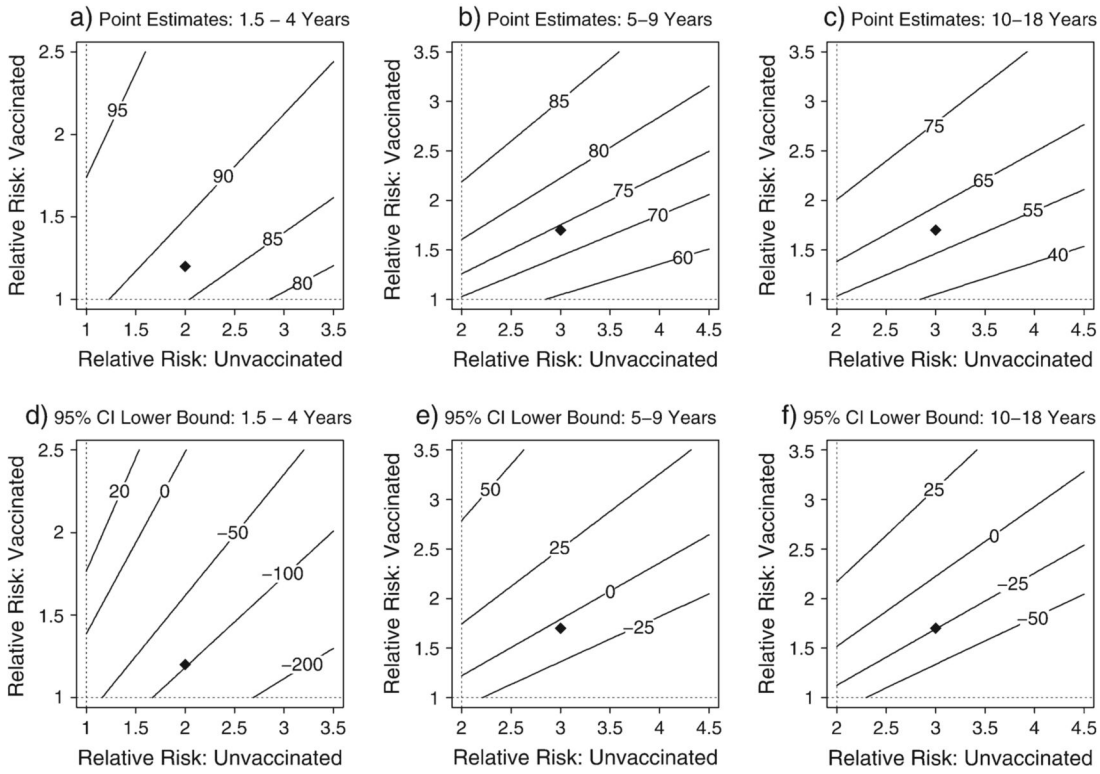

Fig. 2.

Frequentist sensitivity analysis of point estimates and lower 95% confidence bounds for the age-group-specific VE as a function of the relative risk selection bias parameters $\beta_{1,x}$ (vaccinated) and $\beta_{0,x}$ (unvaccinated) varied over the 90% ranges elicited from the expert. Black diamonds indicate the results at the best guess of the expert. Black lines with numbers indicate the contours.

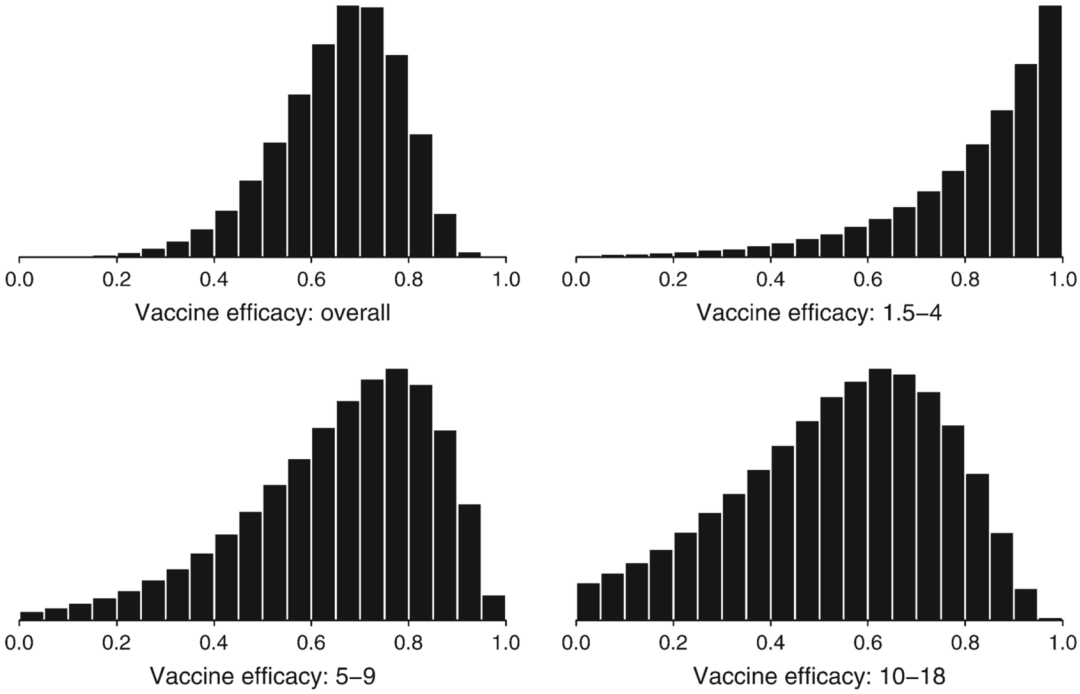

Fig. 3.
Posterior distributions of the overall VE and by age group using the informative prior distributions.

Table 1

Study data for influenza epidemic season 2000–2001 (from Halloran *and others*, 2003)

Age (years)	Vaccine status	Children	MAARI cases	MAARI proportion	MAARI cases cultured	NumberFraction positivecultures culturespositive	Fraction cultured
1.5–4	CAIV-T	537	389	0.72	16	00	0.041
	None	1844	1665	0.90	86	240.28	0.052
5–9	CAIV-T	807	316	0.39	17	20.12	0.054
	None	2232	1156	0.52	118	530.45	0.102
10–18	CAIV-T	937	219	0.23	19	30.16	0.087
	None	5249	1421	0.27	123	560.46	0.087
Total	CAIV-T	2281	924	0.41	52	50.10	0.056
	None	9325	4242	0.45	327	1330.41	0.077

Table 2

Best guess and 90% range for the informative prior distributions on the selection bias parameter **β** and log **β**

Age group (years)		β (relative risk) scale				log β scale			
		Unvaccinated		Vaccinated		Unvaccinated		Vaccinated	
		Best guess	90% range	Best guess	90% range	Best guess	90% range	Best guess	90% range
1.5–4	elicited	2.00	1.00, 3.50	1.20	1.00, 2.50	0.69	0.00, 1.25	0.18	0.00, 0.92
	used	2.00	1.00, 4.00	1.20	0.58, 2.50	0.69	0.00, 1.39	0.18	−0.55, 0.92
5–18	elicited	3.00	2.00, 4.50	1.60	1.00, 3.50	1.10	0.69, 1.50	0.47	0.00, 1.25
	used	3.00	2.00, 4.50	1.70	1.00, 2.89	1.10	0.69, 1.50	0.53	0.00, 1.06

Table 3

Results of Bayesian and frequentist analyses using surveilance cultures. For the Bayesian analyses, the posterior means (95% highest posterior density credible intervals) for vaccine efficacy are reported, for the frequentist analyses, the point estimates (95% confidence intervals). The estimates using just the non-specific MAARI case definition are included for comparison

Analysis	Age group			
	1.5–4 years	**5–9 years**	**10–18 years**	**Overall**
Bayesian				
Informative $\pi(\beta)$	0.80 (0.23, 1.00)	0.65 (0.13, 0.93)	0.51 (−0.12, 0.88)	0.65 (0.35, 0.86)
$\pi(\beta)$ fixed at best guess	0.77 (0.11, 0.99)	0.63 (0.10, 0.93)	0.50 (−0.12, 0.86)	0.64 (0.32, 0.85)
$\pi(\beta)$ fixed at 1 (MAR)	0.84 (0.41, 0.90)	0.73 (0.40, 0.94)	0.64 (0.26, 0.89)	0.73 (0.53, 0.88)
Frequentist				
β fixed at best guess	0.88 (−0.97, 0.99)	0.74 (−0.05, 0.88)	0.61 (−0.25, 0.88)	0.73 (0.34, 0.89)
β fixed at 1 (MAR)	0.91 (−0.34, 0.99)	0.80 (0.26, 0.95)	0.70 (0.13, 0.90)	0.79 (0.52, 0.91)
MAARI alone	0.20 (0.14, 0.25)	0.25 (0.15, 0.34)	0.14 (0.01, 0.26)	0.18 (0.11, 0.24)