Bias reduction for risk ratio and vaccine effect estimators

Stephen E. Chick[1,*,†], Daniel C. Barth-Jones[2] and James S. Koopman[3]

[1]*Department of Industrial and Operations Engineering, University of Michigan, 1205 Beal Avenue, Ann Arbor, MI 48109-2117, U.S.A.*
[2]*Center for Healthcare Effectiveness Research, Wayne State University School of Medicine, 121 Shiffman Medical Library, 4325 Brush St., Detroit, MI 48201, U.S.A.*
[3]*Department of Epidemiology, University of Michigan, 109 Observatory St., Ann Arbor, MI 48109-2029, U.S.A.*

SUMMARY

We examine the structural bias for established estimators of vaccine effects on susceptibility and for newer estimates of vaccine effects on infectiousness. We then propose and analyse new bias corrections for vaccine effect estimators of both susceptibility and infectiousness, as well as their combined effect on infection transmission. Each estimator is evaluated empirically with computer simulations. Of the estimators examined in this paper, those with the least bias and root mean squared error are computed by adding one to the positive count in the placebo population. We also identify a source of bias for a standard Bayesian estimator of risk ratios. Copyright © 2001 John Wiley & Sons, Ltd.

1. INTRODUCTION

Vaccination programmes have proven to be an effective intervention for many infectious diseases. One important vaccine effect is to prevent the reproduction of a infectious agent so that a vaccinated individual is no longer susceptible to infection, regardless of how many times exposure occurs. This is a model II vaccine effect on susceptibility, in the nomenclature of Smith *et al.* [1], that imparts complete immunity to infection for a fraction of vaccine recipients. However, there are several other vaccine effects that may also reduce the transmission of infectious agents in a population as well. The induced immune response may keep agents from reaching a niche where they can reproduce in a way that reduces the probability infection *per exposure* [2, 3], known as a model I [1] vaccine effect on susceptibility. Alternatively, the vaccine induced immune response can slow agent reproduction, lower the level of agent generated, and thus decrease the contagiousness of the host. This reduces the probability that

---

*Correspondence to: Stephen E. Chick, Department of Industrial and Operations Engineering, University of Michigan, 1205 Beal Avenue, Room 1777, Ann Arbor, MI 48109-2117, U.S.A.
† E-mail: sechick@engin.umich.edu

a vaccinated, infected individual may transmit infection to others, an *infectiousness* effect that reduces the rate of secondary infections. The oral polio vaccine, *H. influenzae* type b vaccine, and the malaria transmission-blocking vaccine demonstrate the importance of measuring these different vaccine effects [4, 5]. Additionally, HIV vaccines may need to rely on infectiousness effects to slow the HIV epidemic [5, 6], as HIV-1 viral load is an important risk factor in heterosexual transmission [7].

There is therefore great interest in obtaining accurate estimates of vaccine effects for susceptibility (VE_S) and infectiousness (VE_I), as well as their combined effect (VE_R) on the basic reproduction number [8] of the infectious agent. Several trial designs have been proposed to estimate those effects using maximum likelihood techniques [5, 6, 9].

Vaccine effect estimators, however, may be biased for a number of reasons, including sample selection bias, observer bias, confounding between vaccination and exposure, population heterogeneity, dependencies between exposure risks and infection experience in trial subjects, lack of specificity in the case definition for disease, and the use of an incorrect statistical model [10–16].

Although the standard vaccine effect estimators are consistent, they are biased because they are non-linear functions of other estimators [17]. This leads to a systematic underestimation of vaccine effects, in expectation, even if other sources of bias are completely eliminated. While the bias may be negligible in phase III trials with many thousands of individuals enrolled in each trial arm, numerical experiments below indicate that the error may be non-negligible when there are under 1000 participants per trial, corresponding to small or intermediate size [18] or phase IIb challenge [19] trials. The bias may also be large when an appropriate subset of the data is used to estimate vaccine effects for a specific cohort. Jewell [20] provides a simple bias correction for relative risk and odds ratio estimators. His result applies directly to model II VE_S estimators, because the model II VE_S is the risk ratio subtracted from 1.

Standard estimators for model I VE_S, VE_I and VE_R also have structural bias, because they are non-linear functions of estimators for risk in the placebo and vaccinated populations. However, the structural bias of those estimators has not yet been studied.

This paper evaluates the structural bias of several vaccine effect estimators. Some known bias corrections for model II susceptibility effects are presented along with new bias corrections for model I VE_S, VE_I and VE_R estimators. Both frequentist (Section 2) and Bayesian (Section 3) estimators are considered. Section 4 evaluates both the bias and root mean squared error (RMSE) of each estimator. Simulations indicate that bias corrections are appropriate when estimating risk ratios or vaccine effects for a cohort with an appropriate subset of the data, for phase IIb, small or intermediate size trials, when the trial produces less data than expected from participants dropping out, or when the number of observed infections is less than anticipated during trial design.

2. FREQUENTIST VACCINE EFFECT ESTIMATORS

This section presumes that data from an appropriate vaccine trial design [4–6, 9] is available to estimate VE_S, VE_I and VE_R. We first describe frequentist statistical issues for both model I and model II effects for susceptibility (VE_S), then discuss VE_I and VE_R. Known bias corrections are reviewed and then new ones are proposed.
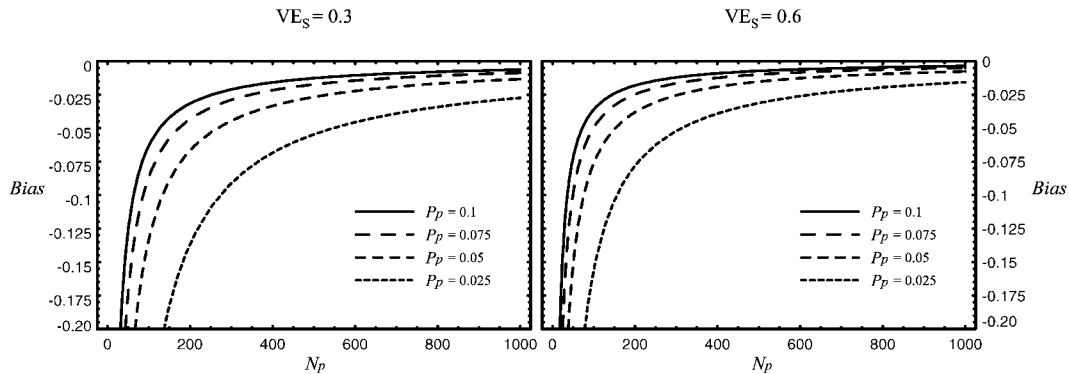
Figure 1. The magnitude of the first-order bias for model II VE_S as a function of the number $n_p = n_v$ of individuals in each arm, the actual vaccine effect ($VE_S = 0.3$ and 0.6), and the probability of infection p_p for a placebo recipient during the trial.

2.1. Model II susceptibility effects

The model II vaccine effect for susceptibility model presumes that a fraction ρ of the vaccinated population is completely protected from all further exposure to infection, and that the remainder of the vaccinated population is unaffected.

Suppose that p_p is the probability of infection over some defined time interval for a placebo recipient, and that p_v is the corresponding probability for a vaccine recipient. The relative risk is $\rho = p_v/p_p$, and the model II VE_S is defined as

$$VE_S \triangleq 1 - p_v/p_p$$

We recall three estimators for VE_S. Let n_p be the size of the uninfected placebo population, and let x_p be the number that become infected during the trial. Let n_v and x_v be the corresponding quantities for the vaccinated population. Assuming conditionally independent Bernoulli trials, Haber $et\ al.$ [2] estimates VE_S as

$$\hat{VE}_S = 1 - \frac{\hat{p}_v}{\hat{p}_p} = 1 - \frac{x_v/n_v}{x_p/n_p} \tag{1}$$

The MLE $\hat{VE}_S$ is consistent, but biased because it is non-linear in the MLE $\hat{p}_p$ for p_p. A standard bias correction $\hat{b}_{S,a.n.}$ can be derived using a Taylor series expansion [17]

$$\hat{\hat{VE}}_S = \hat{VE}_S + \hat{b}_{S,a.n.}$$

$$= \hat{VE}_S + \frac{(1 - \hat{p}_p)\hat{p}_v}{n_p \hat{p}_p^2} \tag{2}$$

Figure 1 illustrates that the magnitude of this first-order bias changes as a function of the number $n_p = n_v$ of individuals per arm, the actual vaccine effect ($VE_S = 0.3$ and 0.6), and the

probability p_p of infection for a placebo recipient during the trial. Doubling either n_p or VE_S halves the bias, and the bias worsens as p_p approaches 0.

Jewell [20] presents a bias correction for the relative risk that is computed as if an extra positive outcome were observed in the placebo population. That correction has an urn model interpretation [21] and determines a third model II VE_S estimator

$$\ddot{\hat{VE}}_S = 1 - \frac{x_v/n_v}{(x_p + 1)/(n_p + 1)} \tag{3}$$

Both bias corrections increase the original estimate.

2.2. Model I susceptibility effects

Model I vaccine effects on susceptibility change the probability of infection per exposure. We recall an existing estimator for model I susceptibility effects, and introduce two new bias corrections. Assume now that the outcome is infection and that infection status is only known at a specified follow-up interval. Further assume that exposure events occur with rate λ for an individual in either the placebo or vaccinated population, and that p_p is the probability of infection *per exposure* in the placebo population. Let p_v be the corresponding probability for the vaccinated population. The model I VE_S is defined to be

$$VE_S \triangleq 1 - p_v/p_p$$

Over an elapsed time τ, the number of exposures has Poisson distribution with mean $\lambda\tau$, so that the probability of infection during the trial is $1 - \exp[-\lambda\tau p_v]$ for vaccinated individuals. The MLE for the attack rate in the vaccinated population is $\hat{AR}_v = x_v/n_v$. Similar results hold for the placebo arm. The MLE $\hat{VE}_S$ for the vaccine effect is determined by substituting the MLE for the attack rates (example, $\hat{AR}_v = 1 - \exp[-\lambda\tau p_v]$)

$$\hat{VE}_S = 1 - \frac{\ln(1 - \hat{AR}_v)}{\ln(1 - \hat{AR}_p)} \tag{4}$$

Haber *et al*. [2] derives this estimator with a differential equation model.

The first new bias correction $\hat{b}_{S,lk}$ for model I susceptibility effects is obtained from the low-order terms of the Taylor series expansion (see the Appendix). This gives the model I effect estimator analogous to the model II effect estimator of equation (2)

$$\begin{aligned}
\hat{\hat{VE}}_S &= \hat{VE}_S + \hat{b}_{S,lk} \\
&= \hat{VE}_S - \frac{\hat{AR}_v}{2n_v(1 - \hat{AR}_v)\ln(1 - \hat{AR}_p)} + \frac{\hat{AR}_p(1 - \hat{VE}_S)(\ln(1 - \hat{AR}_p) + 2)}{2n_p(1 - \hat{AR}_p)(\ln(1 - \hat{AR}_p))^2}
\end{aligned} \tag{5}$$

The second new bias correction for model I vaccine effects for susceptibility is obtained by estimating the attack rate estimate in the placebo population by adding one to the positive count in the placebo population, as with the estimator in equation (3)

$$\ddot{\hat{VE}}_S = 1 - \frac{\ln(1 - x_v/n_v)}{\ln(1 - (x_p + 1)/(n_p + 1))} \tag{6}$$

2.3. Vaccine effects on infectiousness

Vaccination may reduce the per-contact transmission probability from an infected individual. Several vaccine trial designs [4–6] measure vaccine effects on infectiousness (VE_I) in terms of the probability of transmission per contact from infected placebo and vaccine recipients to uninfected, unvaccinated individuals. Let π_p be the transmission probability from an infected placebo recipient to an uninfected, unvaccinated individual, and let π_v be the corresponding probability from an infected vaccine recipient. Then

$$VE_I \triangleq 1 - \pi_v/\pi_p$$

If the number of transmission opportunities during a given partnership between a vaccinated infected individual and an unvaccinated uninfected individual is unknown and assumed to be a Poisson random variable whose mean is the same in both trial arms, VE_I can be estimated as with VE_S in Section 2.2. Section 5 discusses potential violations of this assumption.

Let m_p be the number of partnerships between infected individuals in the placebo arm and uninfected, unvaccinated individuals, and let y_p be the number of transmissions due to those partnerships. In a generic vaccine trial for sexually transmitted infections, m_p would represent the sum of the number of partners reported by each individual in the placebo population. Let m_v and y_v be analogous quantities for the vaccine arm. Then $\hat{SAR}_p = y_p/m_p$ and $\hat{SAR}_v = y_v/m_v$ are the MLEs for the partnership secondary attack rates in the placebo and vaccine arms. This secondary attack rate refers to a *per partnership* probability for secondary infection, rather than a per contact transmission probability. A simple modification of the standard model I VE_S estimator leads to the following estimator for VE_I effects:

$$\hat{VE}_I = 1 - \frac{\ln(1 - \hat{SAR}_v)}{\ln(1 - \hat{SAR}_p)} = 1 - \frac{\ln(1 - y_v/m_v)}{\ln(1 - y_p/m_p)} \tag{7}$$

The first-order bias correction leads to the following new estimator for VE_I:

$$\begin{aligned} \hat{\hat{VE}}_I &= \hat{VE}_I + \hat{b}_{I,lk} \\ &= \hat{VE}_I - \frac{\hat{SAR}_v}{2m_v(1 - \hat{SAR}_v)\ln(1 - \hat{SAR}_p)} + \frac{\hat{SAR}_p(1 - \hat{VE}_I)(\ln(1 - \hat{SAR}_p) + 2)}{2m_p(1 - \hat{SAR}_p)(\ln(1 - \hat{SAR}_p))^2} \end{aligned} \tag{8}$$

A second new bias-corrected VE_I estimator has a secondary attack rate estimator for the placebo population that is analogous to the model I VE_S estimator in equation (6)

$$\ddot{VE}_I = 1 - \frac{\ln(1 - y_v/m_v)}{\ln(1 - (y_p + 1)/(m_p + 1))} \tag{9}$$

2.4. Combined vaccine effects

The combined effect VE_R of vaccination on the basic reproduction number R_0 depends on both VE_S and VE_I

$$VE_R \triangleq 1 - (1 - VE_S)(1 - VE_I)$$

In a homogeneously mixing population, R_0 is interpreted as the expected number of new infections caused by an infected individual in an otherwise entirely susceptible population,

and VE_R gives the relative reduction in R_0 when everybody is vaccinated. When $R_0 > 1$, non-zero endemic infection levels are observed, and when $R_0 < 1$ infections tend to die out. Longini *et al.* [4] indicate that the critical fraction f^* of a homogeneously mixing population that must be vaccinated in order to eliminate endemic infection by reducing the basic reproduction number R_0 below 1 is $f^* = (R_0 - 1)/(R_0 VE_R)$.

One estimator for VE_R is determined by the MLEs for VE_S and VE_I [9, 22]

$$\hat{VE}_R = 1 - \frac{\ln(1 - \hat{AR}_v)}{\ln(1 - \hat{AR}_p)}\frac{\ln(1 - \hat{SAR}_v)}{\ln(1 - \hat{SAR}_p)} \tag{10}$$

Two new bias-corrected VE_R estimators are obtained as above with the Taylor expansion

$$\hat{\hat{VE}}_R = \hat{VE}_R + \hat{b}_{S,lk}\frac{\ln(1 - \hat{SAR}_v)}{\ln(1 - \hat{SAR}_p)} + \frac{\ln(1 - \hat{AR}_v)}{\ln(1 - \hat{AR}_p)}\hat{b}_{I,lk} \tag{11}$$

and with the modified attack rate estimators for the placebo population

$$\ddot{VE}_R = 1 - \frac{\ln(1 - x_v/n_v)}{\ln(1 - (x_p + 1)/(n_p + 1))}\frac{\ln(1 - y_v/m_v)}{\ln(1 - (y_p + 1)/(m_p + 1))} \tag{12}$$

The biology of a particular vaccine might suggest that the model II VE_S estimator is more appropriate for the estimation of VE_R than the model I estimator considered here. The appropriate bias corrections for that vaccine would be derived similarly.

3. BAYESIAN ESTIMATORS

An alternative formulation that is consistent with existing Bayesian biostatistical literature [23–25] is to consider p_p and p_v to be independent random variables whose realizations are to be inferred from vaccine trial data and Bayes rule.

3.1. Bayesian vaccine effects formulation

First consider model II susceptibility effects. Although any joint prior distribution for p_p and p_v is valid, we use independent beta prior distributions for analytical tractability and consistency with Ashby *et al.* [25]

$$\Pr(p_p) \sim beta(\alpha_p, \beta_p) \quad \text{and} \quad \Pr(p_v) \sim beta(\alpha_v, \beta_v) \tag{13}$$

The prior probability density function of p_p on the interval $[0, 1]$ is therefore

$$f(p_p \mid \alpha_p, \beta_p) = \frac{\Gamma(\alpha_p + \beta_p)}{\Gamma(\alpha_p)\Gamma(\beta_p)} p_p^{\alpha_p - 1}(1 - p_p)^{\beta_p - 1}$$

The prior mean is $\alpha_p/(\alpha_p + \beta_p)$ (if $\alpha_p, \beta_p > 0$) and prior mode is $(\alpha_p - 1)/(\alpha_p + \beta_p - 2)$ (if $\alpha_p, \beta_p > 1$). Let $\mathscr{E} = (x_p, n_p, x_v, n_v; y_p, m_p, y_v, m_v)$ summarize the vaccine trial outcomes for the primary and secondary transmissions. The posterior distributions for p_p is

$$\Pr(p_p \mid \mathscr{E}) \sim beta(\alpha_p + x_p, \beta_p + n_p - x_p) \tag{14}$$

The posterior mean $\tilde{p}_p$ and variance $\tilde{\sigma}_p^2$ of the attack rate in the placebo population are

$$\tilde{p}_p = \frac{\alpha_p + x_p}{\alpha_p + \beta_p + n_p}$$

$$\tilde{\sigma}_p^2 = \frac{(\alpha_p + x_p)(\beta_p + n_p - x_p)}{(\alpha_p + \beta_p + n_p)^2(\alpha_p + \beta_p + n_p + 1)} = \frac{\tilde{p}_p(1 - \tilde{p}_p)}{\alpha_p + \beta_p + n_p + 1} \tag{15}$$

Similar formulae hold for the posterior distribution of p_v, and the quantities $\tilde{p}_v$ and $\tilde{\sigma}_v^2$.

For model I susceptibility effects, we use a beta prior distribution for the unknown attack rates AR_p, AR_v, rather than the per-contact transmission probabilities. Similarly, the partnership secondary attack rates SAR_p, SAR_v are presumed to have beta prior distributions. The posterior means and variances of these unknown quantities are computed as in equation (15).

In what follows, we use $(\alpha_p, \beta_p) = (\alpha_v, \beta_v) = (1/2, 1/2)$, a U-shaped, non-informative prior distribution, as did Ashby *et al.* [25] in a cancer risk study. Other choices are (1,1) for a uniform distribution, or to formally let $\alpha, \beta \to 0$ to 'wash out' the influence of the prior.

3.2. Posterior mean for vaccine effects

The posterior expectation, conditional on the information $\mathscr{E}$ obtained from the trial, is an important Bayesian point estimator because it minimizes the squared error loss. The expected model II susceptibility effect

$$E[VE_S \,|\, \mathscr{E}] = E[1 - p_v/p_p \,|\, \mathscr{E}] = 1 - \tilde{p}_v \frac{n_p}{x_p - 0.5} \tag{16}$$

is determined from the distributions in equation (14), assuming that $x_p > 0.5$. The concept of bias correction is not applied in the standard Bayesian literature, so we do not consider a bias correction for $E[VE_S \,|\, \mathscr{E}]$.

Similarly, the posterior expectation of the VE_I and VE_R are

$$E[VE_I \,|\, \mathscr{E}] = E\left[1 - \frac{\ln(1 - \pi_p)}{\ln(1 - \pi_p)}\,\bigg|\, \mathscr{E}\right]$$

$$E[VE_R \,|\, \mathscr{E}] = E\left[1 - \frac{\ln(1 - p_v)}{\ln(1 - p_p)}\frac{\ln(1 - \pi_v)}{\ln(1 - \pi_p)}\,\bigg|\, \mathscr{E}\right]$$

where the posterior distributions for the partnership secondary attack rates π_p, π_v are analogous to equation (14). Monte Carlo sampling can be used to estimate $E[VE_I \,|\, \mathscr{E}], E[VE_R \,|\, \mathscr{E}]$ or to obtain credible sets (Bayesian confidence intervals).

3.3. Bayesian analogues of frequentist estimators

A second class of Bayesian point estimators are structurally similar to the frequentist estimators in Section 2, but substitute Bayesian concepts for frequentist ones. One Bayesian analogue $\tilde{VE}_S$ of the frequentist model II effect estimator $\hat{VE}_S$ uses posterior means instead of MLEs for the attack rate estimates

$$\tilde{VE}_S = 1 - \frac{\tilde{p}_v}{\tilde{p}_p} \tag{17}$$

A Taylor series expansion for $\tilde{VE}_S$ (see the Appendix) motivates the bias correction $\tilde{b}_{S,a.n.}$

$$
\begin{aligned}
\tilde{\tilde{VE}}_S &= \tilde{VE}_S + \tilde{b}_{S,a.n.} \\
&= \tilde{VE}_S + \tilde{\sigma}_p^2 \frac{\tilde{p}_v}{n_p \tilde{p}_p^3} = \tilde{VE}_S + \frac{(1-\tilde{p}_p)\tilde{p}_v}{(n_p+2)\tilde{p}_p^2}
\end{aligned}
\tag{18}
$$

A similar analysis applies for estimators of model I VE_S and VE_I, and of VE_R

$$
\tilde{VE}_S = 1 - \frac{\ln(1-\tilde{p}_v)}{\ln(1-\tilde{p}_p)}
\tag{19}
$$

The Taylor series bias correction gives the estimator

$$
\begin{aligned}
\tilde{\tilde{VE}}_S &= \tilde{VE}_S + \tilde{b}_{S,lk} \\
&= \tilde{VE}_S - \frac{\tilde{\sigma}_v^2}{2(1-\tilde{p}_v)^2 \ln(1-\tilde{p}_p)} + \frac{\tilde{\sigma}_p^2 \ln(1-\tilde{p}_v)}{2(1-\tilde{p}_p)^2(\ln(1-\tilde{p}_p))^2}\left[1 + \frac{2}{\ln(1-\tilde{p}_p)}\right]
\end{aligned}
\tag{20}
$$

For the corresponding VE_I estimators $\tilde{VE}_I$ and $\tilde{\tilde{VE}}_I = \tilde{VE}_I + \tilde{b}_{I,lk}$, replace the attack rate estimators $\tilde{p}_p, \tilde{p}_v, \tilde{\sigma}_p^2, \tilde{\sigma}_v^2$ with appropriate secondary attack rate estimators. For VE_R

$$
\tilde{VE}_R = 1 - (1-\tilde{VE}_S)(1-\tilde{VE}_I)
\tag{21}
$$

and the Taylor series expansion bias correction gives

$$
\tilde{\tilde{VE}}_R = \tilde{VE}_R + \tilde{b}_{S,lk}(1-\tilde{VE}_I) + (1-\tilde{VE}_S)\tilde{b}_{I,lk}
\tag{22}
$$

The MLE $\hat{VE}_S$ in equation (1) can be formally written $1 - E[p_v \mid \mathscr{E}]/E[p_p \mid \mathscr{E}]$ in the limit as $\alpha_p, \beta_p, \alpha_v, \beta_v \to 0$, and $\ddot{VE}_S$ can be similarly expressed when $\alpha_p = 1$ and $\beta_p, \alpha_v, \beta_v \to 0$. Jewell [20] notes that the estimator $1 - \ddot{VE}_S$ is not invariant to labelling changes (exchange placebo with vaccine arms). The estimator $1 - \tilde{VE}_S$ *is* invariant to such changes.

4. EMPIRICAL EVALUATION OF VACCINE EFFECT ESTIMATORS

The bias and root mean squared error (RMSE) of all the above estimators are estimated empirically over a range of trial population sizes and known vaccine effects. The goal is to identify the situations where the various estimators for each vaccine effect perform particularly well or poorly in terms of bias and RMSE.

The placebo and vaccine population sizes, n_p, n_v, were set to 100, 300 and 1000 to evaluate the effect of study population size on the bias and RMSE. These numbers correspond to small or intermediate size trials. While a reduction in susceptibility to infection of at least 60 per cent might be considered to be a substantial effect, an intermediate size trial might be designed

Table I. Bias estimates for model II VE_S estimators (identified by bias correction). The symbol * indicates that trials with undefined estimators (division by 0) were ignored in the computation.

Set-up		Frequentist			Bayesian		
		MLE	Taylor series	Jewell	Mean	(none)	Taylor series
$n_v = n_p$	VE_S	$\hat{VE}_S$	$\hat{\hat{VE}}_S$	$\ddot{VE}_S$	$E[VE_S \mid \mathscr{E}]$	$\tilde{VE}_S$	$\tilde{\tilde{VE}}_S$
100	0.00	-0.190^*	0.066^*	-0.013	-0.421	-0.180	0.060
	0.30	-0.153^*	0.038^*	-0.009	-0.344	-0.145	0.007
	0.60	-0.066^*	0.031^*	0.009	-0.201	-0.098	-0.013
300	0.00	-0.043	0.007	0.005	-0.097	-0.041	0.008
	0.30	-0.038	-0.001	-0.003	-0.068	-0.043	-0.008
	0.60	-0.114	0.007	0.006	-0.050	-0.027	-0.007
1000	0.00	-0.007	0.007	0.007	-0.031	-0.006	0.007
	0.30	-0.011	-0.002	-0.002	-0.028	-0.014	-0.004
	0.60	-0.006	0.000	0.000	-0.014	-0.010	-0.004

to screen out candidate vaccines with observed efficacies of 30 per cent or less [18], and there may justification for studying low-efficacy vaccines [26]. The simulated susceptibility and infectiousness effects are therefore set to 0.0, 0.3 and 0.6 to see how the bias changes with the magnitude of the effect. The study focuses on the bias and RMSE of the vaccine estimates, and does not consider issues related to confidence intervals for standard estimators that have been treated elsewhere [27].

For each population size and vaccine effect, estimates are generated from the sample statistics of 4000 independent trials. In each trial, the simulated vaccine effect reflects the assumptions of the estimator (for example, model I effects are simulated for model I estimators).

For the model II VE_S simulations, the probability of infection in the placebo population is 0.07, and the probability of infection in the vaccinated population is $0.07 \times (1 - VE_S)$. Table I presents bias estimates for all six all-or-one VE_S estimators, and the RMSE is presented in Table II. The standard error ranges from $0.006 - 0.013$ when $n_p = n_v = 100$, and is about $0.0013 - 0.0025$ when $n_p = n_v = 1000$.

To aid in reading the tables, we recall that a hat $\hat{}$ denotes an MLE, and a tilde $\tilde{}$ indicates that Bayesian point estimates of the individual attack rates are used. Two symbols ($\hat{\hat{}}$ or $\tilde{\tilde{}}$) denote a first-order bias corrected estimator. The symbol $\ddot{}$ indicates that the attack rate for the placebo population is estimated by adding 1 to the number of positives. Finally, $E[VE \mid \mathscr{E}]$ represents the Bayesian posterior expectation of a vaccine effect.

The conclusions drawn from the model II susceptibility estimators are equally applicable to relative risk estimators in general, since they are linear functions of each other. The estimators $\ddot{VE}_S$ and $\tilde{\tilde{VE}}_S$ perform better than the other four estimators with respect to the magnitude of the bias. Their performance is within error for most situations, although $\ddot{VE}_S$ appears somewhat better for small populations ($n_p = n_v = 100$). The first-order bias corrections from the Taylor series expansion improves the bias of the estimators ($\hat{\hat{VE}}_S$ has less bias than $\hat{VE}_S$, and $\tilde{\tilde{VE}}_S$ has less bias than $\tilde{VE}_S$). The estimators $\hat{VE}_S$ and $\hat{\hat{VE}}_S$ are undefined when the number of

Table II. Root mean squared error (RMSE) estimates for model II VE_S estimators. The symbol $*$ indicates that trials with undefined estimators (division by 0) were ignored in the computation.

Set-up		Frequentist			Bayesian		
$n_v = n_p$	VE_S	$\hat{VE}_S$	$\hat{\hat{VE}}_S$	$\ddot{VE}_S$	$E[VE_S \mid \mathscr{E}]$	$\tilde{VE}_S$	$\tilde{\tilde{VE}}_S$
100	0.00	0.860*	0.459*	0.637	1.502	0.973	0.864
	0.30	0.695*	0.365*	0.447	1.230	0.574	0.354
	0.60	0.389*	0.258*	0.284	0.639	0.356	0.247
300	0.00	0.345	0.314	0.316	0.357	0.334	0.305
	0.30	0.263	0.240	0.242	0.278	0.258	0.235
	0.60	0.174	0.162	0.163	0.185	0.173	0.159
1000	0.00	0.166	0.163	0.163	0.168	0.165	0.161
	0.30	0.131	0.128	0.128	0.135	0.130	0.127
	0.60	0.087	0.085	0.085	0.093	0.087	0.084

positive individuals in the placebo population is 0. The output of trials where that occurs is ignored, and results based sample averages where some trials are ignored are marked with $*$.

The posterior expectation $E[VE_S \mid \mathscr{E}]$ of the vaccine effect has the worst bias of the six estimators considered here, and that the bias is significant even with $n_p = n_v = 300$ individuals per trial arm. In fact, it has twice the low-order bias of $\tilde{VE}_S$ because $E[VE_S \mid \mathscr{E}] \approx \tilde{VE}_S - \tilde{b}_{S,\text{a.n.}}$, whereas $\tilde{\tilde{VE}}_S = \tilde{VE}_S + \tilde{b}_{S,\text{a.n.}}$ is a low-bias estimator.

We conjecture that the prior distribution in equation (13) for the unknown probabilities p_p, p_v may be a culprit. In particular, one reason for implementing a vaccine trial is an *a priori* belief that $p_v < p_p$, although there may be no effect ($p_v = p_p$) with a non-zero probability. Equation (13) implies, somewhat contradictorily, that the two unknown probabilities are statistically independent and there is 0 probability for no vaccine effect. This suggests that using non-informative prior distributions for vaccine effects specifically, and for risk ratios in general, may lead to unintended and undesirable consequences. There appears to be a need for further work to help decision-makers specify prior distributions for vaccine effect estimates that account for the beliefs about the likely biological effects of vaccines.

Table II illustrates that the RMSE of all of the model II VE_S estimators (except $E[VE_S \mid \mathscr{E}]$) is roughly the same for larger population sizes, although the estimators with a Taylor series bias correction ($\hat{\hat{VE}}_S$ and $\tilde{\tilde{VE}}_S$) seem to have a slight advantage over $\ddot{VE}_S$, which is in turn somewhat better than the others. That the RMSE for $\hat{\hat{VE}}_S$ is particularly good when $n_p = n_v = 100$ is an artifact of the fact that trials with undefined estimators (division by 0) were not included in the computation.

The simulations that evaluate the model I VE_S estimators assume that each trial participant experiences on average 10 exposures, with a Poisson distribution for the actual number of exposures. The per-contact transmission probability is set to $\ln(1 - 0.07)/10 \approx 0.00726$ in the placebo arm to obtain an attack rate of 0.07 in the placebo arm. The per-contact transmission rate in the vaccinated population is $(1 - VE_S)\ln(1 - 0.07)/10$. The estimator with the smallest bias is again $\ddot{VE}_S$, and the relative performance of the other estimators parallels that for the model II estimators, as shown in Table III.

Table III. Bias estimates for model I VE_S estimators. The symbol * indicates that trials with undefined estimators (division by 0) were ignored in the computation.

Set-up		Frequentist			Bayesian		
$n_v = n_p$	VE_S	$\hat{VE}_S$	$\hat{\hat{VE}}_S$	$\ddot{\hat{VE}}_S$	$E[VE_S\,|\,\mathscr{E}]$	$\tilde{VE}_S$	$\tilde{\tilde{VE}}_S$
100	0.00	-0.216^*	0.057^*	0.002	-0.904	-0.170	0.086
	0.30	-0.139^*	0.051^*	-0.005	-0.477	-0.161	0.022
	0.60	-0.081^*	0.028^*	-0.004	-0.327	-0.118	-0.012
300	0.00	-0.051	0.004	-0.001	-0.107	-0.049	0.005
	0.30	-0.044	-0.005	-0.008	-0.094	-0.049	-0.011
	0.60	-0.025	-0.003	-0.005	-0.061	-0.038	-0.016
1000	0.00	-0.010	0.005	0.004	-0.023	-0.010	0.005
	0.30	-0.011	-0.001	-0.001	-0.026	-0.013	-0.003
	0.60	-0.007	-0.001	-0.002	-0.015	-0.011	-0.005

Table IV. Bias estimates for VE_I estimators. The symbol * indicates that trials with undefined estimators (division by 0) were ignored in the computation. Results assume that $VE_S = 0.0$ and that each index case has a mean of 10 partners.

Set-up		Frequentist			Bayesian		
$n_v = n_p$	VE_I	$\hat{VE}_I$	$\hat{\hat{VE}}_I$	$\ddot{\hat{VE}}_I$	$E[VE_I\,|\,\mathscr{E}]$	$\tilde{VE}_I$	$\tilde{\tilde{VE}}_I$
100	0.00	-0.149^*	0.047^*	-0.001	-0.322	-0.124	0.034
	0.30	-0.094^*	0.030^*	0.002	-0.255	-0.098	0.001
	0.60	-0.056^*	0.026^*	0.005	-0.171	-0.080	-0.010
300	0.00	-0.031	0.008	0.004	-0.069	-0.030	0.008
	0.30	-0.025	0.002	0.000	-0.051	-0.029	-0.003
	0.60	-0.016	0.000	0.002	-0.045	-0.025	-0.010
1000	0.00	-0.011	-0.000	-0.001	-0.019	-0.011	-0.000
	0.30	-0.008	0.000	0.000	-0.014	-0.009	-0.002
	0.60	-0.006	-0.002	-0.002	-0.014	-0.009	-0.005

Table IV summarizes the bias of the model I VE_I estimators. The number of exposures per susceptible person is assumed to be a Poisson random variable with mean 10. The per-contact transmission probability is $\ln(1-0.10)/10$ so that roughly 10 per cent of the initially susceptible population becomes infected. Each new infectee contacts a random number (Poisson with mean 10) of unvaccinated susceptibles, and exposes each contact a random number (Poisson with mean 10) of times. For these simulations, $VE_S = 0$. The random number of outcomes for secondary infections (m_p, m_v) does not seem to change the relative ordering of the estimators in terms of their bias characteristics.

The RMSE for the model I VE_S and VE_I estimators (data not shown) are similar to those for the model II VE_S estimators, in that estimators with a Taylor series bias correction have a slight advantage over $\ddot{\hat{VE}}_S$, which is in turn somewhat better than the others.

Table V. Bias estimates for VE_R estimators.

Set-up				Frequentist			Bayesian		
$n_v = n_p$	VE_S	VE_I	VE_R	$\hat{VE}_R$	$\hat{\hat{VE}}_R$	$\ddot{VE}_R$	$E[VE_R \mid \mathscr{E}]$	$\tilde{VE}_R$	$\tilde{\tilde{VE}}_R$
300	0.00	0.00	0.00	−0.074	−0.067	−0.001	−0.115	−0.070	−0.060
		0.30	0.30	−0.055	−0.050	−0.004	−0.087	−0.057	−0.050
		0.60	0.60	−0.032	−0.029	−0.003	−0.055	−0.040	−0.036
	0.30	0.00	0.30	−0.055	−0.050	−0.004	−0.099	−0.061	−0.063
		0.30	0.51	−0.040	−0.037	−0.005	−0.070	−0.050	−0.052
		0.60	0.72	−0.023	−0.021	−0.002	−0.049	−0.035	−0.036
	0.60	0.00	0.60	−0.033	−0.029	−0.004	−0.073	−0.050	−0.064
		0.30	0.72	−0.022	−0.099	−0.002	−0.057	−0.039	−0.049
		0.60	0.86	−0.034	−0.033	−0.023	−0.055	−0.051	−0.057
1000	0.30	0.30	0.51	−0.011	−0.011	−0.002	−0.021	−0.015	−0.016
		0.60	0.72	−0.008	−0.008	−0.003	−0.015	−0.013	−0.013
	0.60	0.30	0.72	−0.008	−0.008	−0.003	−0.018	−0.014	−0.017
		0.60	0.86	−0.022	−0.022	−0.019	−0.028	−0.027	−0.029

Table VI. Estimate of the RMSE of VE_R estimators.

Set-up				Frequentist			Bayesian		
$n_v = n_p$	VE_S	VE_I	VE_R	$\hat{VE}_R$	$\hat{\hat{VE}}_R$	$\ddot{VE}_R$	$E[VE_R \mid \mathscr{E}]$	$\tilde{VE}_R$	$\tilde{\tilde{VE}}_R$
300	0.00	0.00	0.00	0.4362	0.4247	0.3873	0.5760	0.4225	0.4032
		0.30	0.30	0.3254	0.3183	0.2895	0.4637	0.3173	0.3041
		0.60	0.60	0.2064	0.2017	0.1841	0.2755	0.2035	0.1952
	0.30	0.00	0.30	0.3329	0.3251	0.2963	0.4465	0.3257	0.3154
		0.30	0.51	0.2494	0.2433	0.2218	0.3312	0.2462	0.2385
		0.60	0.72	0.1559	0.1526	0.1404	0.2302	0.1565	0.1533
	0.60	0.00	0.60	0.2336	0.2305	0.2141	0.3783	0.2337	0.2348
		0.30	0.72	0.1703	0.1673	0.1543	0.2510	0.1726	0.1744
		0.60	0.86	0.1105	0.1086	0.1005	0.1556	0.1145	0.1162
1000	0.30	0.30	0.51	0.1162	0.1158	0.1131	0.1456	0.1162	0.1158
		0.60	0.72	0.0775	0.0768	0.0748	0.1020	0.0775	0.0775
	0.60	0.30	0.72	0.0806	0.0800	0.0781	0.1044	0.0812	0.0819
		0.60	0.86	0.0539	0.0539	0.0529	0.0721	0.0548	0.0548

The simulated trials for VE_R effects use the same assumptions used for the VE_I simulations, except that both VE_S and VE_I are varied. Adjusting the bias of the attack rates by 'adding' a positive observation to the placebo population ($\ddot{VE}_R$) again gives the best bias characteristics for the vaccine settings considered here, as supported by the bias estimates given in Table V. Unlike the other studies, however, the low-order Taylor series bias correction does not seem to appreciably improve the bias of the estimators. Table VI indicates that $\ddot{VE}_R$ exhibits the lowest RMSE. For a given VE_R, the RMSE is slightly but consistently larger when VE_S

is larger. If $n_p = n_v = 1000$, for example, the RMSE of $\ddot{\mathrm{VE}}_R$ is 0.170 when $\mathrm{VE}_S = 0.6$ and $\mathrm{VE}_I = 0.3$, but is 0.155 when $\mathrm{VE}_S = 0.3$ and $\mathrm{VE}_I = 0.6$.

The bias and RMSE of each estimator generally decrease as either or both of VE_S and VE_I change from 0.0 to 0.3 to 0.6. The notable exception is a moderately increased bias when $\mathrm{VE}_S = \mathrm{VE}_I = 0.6$, so that $\mathrm{VE}_R = 0.84$. When vaccine effects are nearly one, the bias increases because fewer secondary infections are observed.

Additional simulations (results not shown) indicate that increasing the probability of becoming infected during the trial from 0.07–0.10 to 0.3 reduces the bias and RMSE, and reducing the probability of becoming infected during the trial to 0.05 increases the bias and RMSE, but the relative performance of the estimators does not appreciably change. When the simulation parameters are modified so that 45 infections are expected in the control arm, as has been suggested for some intermediate-size trials [18], the bias is somewhat improved with the Jewell-like estimators, but there is not a large improvement in the RMSE. These shifts are fully anticipated, as a somewhat increased probability of becoming infected during the trial leads to a smaller relative error in the attack rate estimators.

5. CONCLUSIONS

There are several vaccine effects of interest in the control of infectious diseases, and many reasons why estimators of vaccine effects may be biased. This study examines the bias that result from the non-linear structure of risk ratios.

Of six estimators for each of a variety of vaccine effect estimators (several permutations of model I or model II effects for susceptibility, infectiousness or combined effects), including several new estimators, numerical results indicate that the estimator with the best bias and RMSE characteristics is computed by adding one to the count of positive outcomes in the placebo population. Although implementing this bias correction might seem to 'corrupt the data', doing so reduces both the bias and RMSE of standard vaccine effect estimators.

The bias corrections above are most valuable when estimating risk ratios or vaccine effects for a cohort with an appropriate subset of data, when the trial size is limited due to economic constraints, when the trial produces less data than expected from participants dropping out, or when the number of observed infections is somewhat less than anticipated during trial design. The structural bias is negligible, however, for large phase III trials.

Our analysis also highlights a systematic bias for a standard Bayesian treatment for risk ratios. The posterior expectation of vaccine effects has twice the low-order bias of the MLE when independent non-informative prior distributions are used for the attack rates in the placebo and treatment populations. We conjecture that less bias might be achieved with prior distributions that better reflect the reasonable belief that a trial is being performed because a treatment has the potential of a beneficial effect, although a potential for no effect exists. Elicitation methods for appropriate prior distributions are an area for future work.

The assumption in Section 2.3 that the number of exposures is a Poisson random variable with common mean may not always be applicable. Partner tracing for sexually transmitted infections may provide more precise contact counts between each partner. In that case, partnership secondary attack rates can be estimated by conditioning on the number of reported contacts between infected participants and susceptible partners [28], then using group test statistics [29]. If the number exposures per partner has a Poisson distribution with known

means that vary in different population subgroups, limiting dilution assay statistics are appropriate [30, 31]. Standard vaccine effect estimators are readily adapted to these situations by using the most appropriate secondary attack rate estimator. Still, reliable contact data may be difficult to obtain. Longini *et al.* [4] therefore provide alternative vaccine efficacy estimators based on data from multiple populations. Bias reductions for vaccine effect estimators based on group tests, limiting dilution assays, and estimators of Longini *et al.* [4] are not proposed here.

In many trials it is important to assess vaccine effects as a function of multiple covariates. The development and analysis of bias corrections for those vaccine effect estimators is an area for future research.

APPENDIX

Low-order bias correction for frequentist estimators in Section 2

Each vaccine effect estimator is a ratio of functions of probabilities subtracted from 1. Let p_p and p_v be probabilities to be estimated from n_p and n_v outcomes of Bernoulli trials, and let

$$\mathrm{VE} = \mathrm{VE}(p_p, p_v) = 1 - \frac{f(p_v)}{g(p_p)}$$

The first few terms of the Taylor series expansion for $\hat{\mathrm{VE}} = \mathrm{VE}(\hat{p}_p, \hat{p}_v)$ about (p_p, p_v) are

$$\hat{\mathrm{VE}} = \mathrm{VE} - (\hat{p}_v - p_v)\frac{f'(\hat{p}_v)}{g(\hat{p}_p)} + (\hat{p}_p - p_p)\frac{f(\hat{p}_v)g'(\hat{p}_p)}{g(\hat{p}_p)^2}$$

$$+ \frac{1}{2}(\hat{p}_v - p_v \ \hat{p}_p - p_p)\begin{pmatrix} \frac{-f''(p_v)}{g(p_p)} & \frac{f'(p_v)g'(p_p)}{g(p_p)^2} \\ \frac{f'(p_v)g'(p_p)}{g(p_p)^2} & \frac{f(p_v)}{g(p_p)^2}[g''(p_p) - \frac{2g'(p_p)^2}{g(p_p)}] \end{pmatrix}(\hat{p}_v - p_v \ \hat{p}_p - p_p)^T + \cdots$$

The low-order bias depends on the matrix of second derivatives

$$E[\hat{\mathrm{VE}}] \approx \mathrm{VE} - \mathrm{var}[\hat{p}_v]\frac{f''(p_v)}{2g(p_p)} + \mathrm{cov}[\hat{p}_v, \hat{p}_p]\frac{f'(p_v)g'(p_p)}{g(p_p)^2}$$

$$+ \mathrm{var}[\hat{p}_p]\frac{f(p_v)}{2g(p_p)^2}\left[g''(p_p) - \frac{2g'(p_p)^2}{g(p_p)}\right]$$

Since infection outcomes are assumed conditionally independent, given p_p, p_v, we have $\mathrm{cov}[\hat{p}_p, \hat{p}_v] = 0$. Since the proportions are estimated with Bernoulli trials, $\mathrm{var}[\hat{p}_p] = p_p(1 - p_p)/n_p$ and $\mathrm{var}[\hat{p}_v] = p_v(1 - p_v)/n_v$. This suggests the first-order bias adjustment

$$\hat{\hat{\mathrm{VE}}} = \hat{\mathrm{VE}} + \frac{\hat{p}_v(1 - \hat{p}_v)f''(\hat{p}_v)}{2n_v g(\hat{p}_p)} - \frac{\hat{p}_p(1 - \hat{p}_p)f(\hat{p}_v)}{2n_p g(\hat{p}_p)^2}\left[g''(\hat{p}_p) - \frac{2g'(\hat{p}_p)^2}{g(\hat{p}_p)}\right]$$

For a model II effect, $f(p_v) = p_v$ and $g(p_p) = p_p$, so that $f' = g' = 1$, and $f'' = g'' = 0$. This leads to the bias correction in equation (2). For a model I effect, $f(p_v) = \ln(1 - p_v)$ so

that $f' = -1/(1 - p_v)$, and $f'' = -1/(1 - p_v)^2$. The derivatives of $g(p_p)$ are similar. This leads to the bias correction in equation (5) and equation (8). The above analysis together with the derivative rule for products of functions lead to the bias correction in equation (11).

Low-order correction for Bayesian point estimators

The low-order bias correction in the Bayesian model II susceptibility estimator in Section 3.3 is derived like the above classical estimators, except that one uses: (a) the Taylor series expansion for $VE(p_p, p_v)$ around the posterior means $(\tilde{p}_p, \tilde{p}_v)$, rather than the expansion for $VE(\hat{p}_p, \hat{p}_v)$ around (p_p, p_v); (b) expectations with respect to the posterior distribution of (p_p, p_v), rather than with respect to the conditional distribution of $\hat{p}_p, \hat{p}_v$, given p_p, p_v; and (c) the posterior variance $\tilde{\sigma}_p^2$ for p_p in equation (15), rather than $var[\hat{p}_p] = p_p(1 - p_p)/n_p$ (similarly for p_v). The other bias corrections in Section 3.3 are handled similarly.

REFERENCES

1. Smith PG, Rodrigues LC, Fine PE. Assessment of the protective efficacy of vaccines against common diseases using case-control and cohort studies. *International Journal of Epidemiology* 1984; **13**(1):87–93.
2. Haber M, Longini IM, Halloran ME. Estimation of vaccine efficacy in outbreaks of infectious diseases. *Statistics in Medicine* 1991; **10**:1573–1584.
3. Koopman JS, Simon CP, Jacquez JA. Assessing contagiousness effects of vaccines and risk factors for transmission. In *Modeling the AIDS Epidemic*: *Planning*, *Policy*, *and Prediction*, Kaplan EH, Brandeau ML (eds). Raven Press: New York, 1994; 439–460.
4. Longini IM, Sagatelian K, Rida WN, Halloran ME. Optimal vaccine trial design when estimating vaccine efficacy for susceptibility and infectiousness from multiple populations. *Statistics in Medicine* 1998; **17**(10):1121–1136.
5. Longini IM, Hudgens MG, Halloran ME. Estimation of vaccine efficacy for both susceptibility to infection and reduction in infectiousness for prophylactic HIV vaccines with partner augmentation. In *The Quantitative Evaluation of HIV Prevention Programs*, Kaplan E, Brookmeyer R (eds). Yale University Press: New Haven, CT, 1999.
6. Koopman JS, Little RJ. Assessing HIV vaccine effects. *American Journal of Epidemiology* 1995; **142**: 1113–1120.
7. Quinn TC, Wawer MJ, Sewankambo N, Serwadda D, Li C, Wabwire-Mangen F, Meehan MO, Lutalo T, Gray RH. Viral load and heterosexual transmission of human immunodeficiency virus type 1. *New England Journal of Medicine* 2000; **342**(13):921–929.
8. Dietz K. The estimation of the basic reproduction number for infectious diseases. *Statistical Methods in Medical Research* 1993; **2**:23–41.
9. Adams AL, Barth-Jones DC, Chick SE, Koopman JS. Simulations to evaluate HIV vaccine trial methods. *Simulation* 1998; **71**(4):228–241.
10. Murray DM. *Design and Analysis of Group-Randomized Trials*. Oxford University Press: New York, 1998.
11. Austin MA, Criqui MH, Barnett-Conner E, Holdbrook MJ. The effect of response bias on the odds ratio. *American Journal of Epidemiology* 1981; **114**(1):137–143.
12. Greenland S. Bias in indirectly adjusted comparisons due to taking the total study population as the reference group. *Statistics in Medicine* 1987; **6**:193–195.
13. Sawyer S. Maximum likelihood estimators for incorrect models, with an application to ascertainment bias for continuous characters. *Theoretical Population Biology* 1990; **38**:351–366.
14. Donnelly P. The correlation structure of epidemic models. *Mathematical Biosciences* 1993; **117**:49–75.
15. Boily M.-C, Mâsse B, Desai K, Alary M, Anderson R. Some important issues in the planning of phase III HIV vaccine efficacy trials. *Vaccine* 1999; **17**:989–1004.

16. Farrington CP. Quantifying misclassification bias in cohort studies of vaccine efficacy. *Statistics in Medicine* 1990; **9**:1327–1337.
17. Rice JA. *Mathematical Statistics and Data Analysis*, 2nd edn. Wadsworth Publishing Company: Belmont, CA, 1995.
18. Rida W, Fast P, Hoff R, Fleming T. Intermediate-size trials for the evaluation of HIV vaccine candidates: A workshop summary. *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology* 1997; **16**:195–203.
19. Tacket CO, Rennels MB, Mattheis MJ. Initial clinical evaluation of new vaccine candidates: phase 1 and 2 clinical trials of safety, immunogenicity, and preliminary efficacy. In *New Generation Vaccines*, 2nd edn, Levine MM, Woodrow GC, Kaper JB, Cobon GS (eds). Marcel Dekker: New York, 1997; 35–45.
20. Jewell NP. On the bias of commonly used measures of association for 2×2 tables. *Biometrics* 1986; **42**: 351–358.
21. Hernández-Suárez CM, Castillo-Chavez C. Urn models and vaccine efficacy estimation. *Statistics in Medicine* 2000; **19**:827–835.
22. Barth-Jones DC. The retrospective partner trials (RPT) HIV vaccine study design for the measurement of vaccine effects on susceptibility and infectiousness. PhD thesis, University of Michigan, Ann Arbor, MI, 1999.
23. Breslow N. Biostatistics and Bayes. *Statistical Science* 1990; **5**(3):269–298.
24. Berry DA, Stangl DK. *Bayesian Biostatistics*. Marcel Dekker: New York, 1996.
25. Ashby D, Hutton JL, McGee MA. Simple Bayesian analysis for case-control studies in cancer epidemiology. *Statistician* 1993; **42**:385–397.
26. Anderson RM, Garnett GP. Low-efficacy HIV vaccines: potential for community-based intervention programmes. *Lancet* 1996; **348**(9033):1010–1013.
27. Ewell M. Comparing methods for calculating confidence intervals for vaccine efficacy. *Statistics in Medicine* 1996; **15**:2379–2392.
28. Halloran ME, Struchiner CJ. Causal inference for infectious diseases. *Epidemiology* 1995; **6**:142–151.
29. Dorfman R. The detection of defective members of a large population. *Annals of Mathematical Statistics* 1943; **14**:436–440.
30. Lefkovits I, Waldemann H. *Limiting Dilution Analysis of Cells in the Immune System*. Cambridge University Press: Cambridge, 1979.
31. Chick SE. Bayesian models for limiting dilution assay and group test data. *Biometrics* 1996; **52**:1055–1062.