STATISTICS IN MEDICINE
Statist. Med. 2006; **25**:1079–1093
Published online 15 November 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/sim.2236

Estimating vaccine effects from studies of outbreaks in household pairs

Niels G. Becker[1,*,†], Tom Britton[2] and Philip D. O'Neill[3]

[1]*National Centre for Epidemiology and Population Health, Australian National University, Canberra ACT 0200, Australia*
[2]*Department of Mathematics, Stockholm University, SE-106 91 Stockholm, Sweden*
[3]*School of Mathematical Sciences, Nottingham University, Nottingham NG7 2RD, U.K.*

SUMMARY

The traditional way to measure efficacy of a vaccine, with respect to reduced susceptibility and reduced infectivity once infected, is to look at relative attack rates. Although straightforward to apply, such measures do not take disease transmission into account, with the consequence that they can depend strongly on the community setting, the duration of the study period, the way participants are recruited into the study and the virulence of the infection. Sometimes they give a very misleading assessment of the vaccine, as we illustrate by examples. Here measures of vaccine efficacy are considered that avoid these defects, and estimation procedures are presented for studies based on outbreaks in household pairs. Such studies enable estimation of vaccine effects on susceptibility, infectivity and transmission. We propose that the vaccine efficacy measures be estimated, without making any assumptions about the nature of the vaccine response, by consistent estimates of bounds for the measures. Copyright © 2005 John Wiley & Sons, Ltd.

KEY WORDS: household pairs study; infectivity; relative attack rates; susceptibility; vaccine efficacy

1. INTRODUCTION

Vaccination is frequently an effective way to control infectious diseases. However, vaccines generally do not achieve the ideal of providing full protection against infection, making it necessary to assess the effectiveness of each vaccine. Two major purposes of such an assessment are to judge (i) the degree of protection it offers an individual against infection, and (ii) its suitability for use in mass vaccination schedules to protect the entire population. The latter requires consideration of the effect of a vaccine on both the susceptibility of vaccinees and the infectivity of vaccinated individuals who happen to get infected.

*Correspondence to: Niels G. Becker, National Centre for Epidemiology and Population Health, Australian National University, Canberra ACT 0200, Australia.
†E-mail: Niels.Becker@anu.edu.au

Received 1 July 2004
Accepted 25 February 2005

Measures of *vaccine efficacy* are made from vaccine trials, where vaccines are stored and administered under optimal conditions, while studies in the field measure *vaccine effectiveness*; see Clements *et al.* [1]. Here we use the term vaccine efficacy although our discussion has relevance to both concepts.

Many studies estimate vaccine efficacy by 1−(relative attack rate); see for example Ornstein *et al.* [2], Chen and Ornstein [3], Halloran *et al.* [4, 5]. We demonstrate, by examples, that estimates of vaccine efficacy based on relative risk can be very misleading, because they do not account for the fact that cases arise as a result of transmission.

These problems motivate us to consider measures of vaccine efficacy that both acknowledge transmission and allow individuals to have a wide range of responses to vaccination. Some progress in this direction was made in Reference [6], which focused on Markov chain Monte Carlo methods to estimate certain efficacy measures associated with the all-partial-none vaccine response model (defined below). The methods are applicable to outbreak data from households of arbitrary size. In contrast, in the present paper we consider a more general vaccine response framework, define a new measure of vaccine effect on infectivity, and restrict attention to methods of inference using data on outbreak size in a sample of household pairs. The attractions of studies based on pairs are that (i) they provide information about the vaccine effect on infectivity and (ii) parameter estimation is more tractable for these studies than they are for studies with larger households. The latter point enables us to make theoretical progress, for example by deriving estimable bounds on vaccine efficacy. Koopman and Little [7], Longini *et al.* [8], Rida [9] and Datta *et al.* [10, 11] also note that studies based on pairs are well suited for the analysis of vaccine efficacy. The present approach differs from previous work by describing the vaccine response differently. We permit a wide range of vaccine responses that leads to concepts of vaccine efficacy having an interpretation that is consistent over study designs. The challenge of estimability within this large family of vaccine responses is overcome by deriving estimable bounds for the efficacy concepts.

The paper is organized as follows. Section 2 describes our models for disease transmission and vaccine response, and proposes efficacy measures. Section 3 shows, via examples, that measures of vaccine efficacy based on relative attack rates can give very misleading assessments of vaccines in settings that arise in practice. Section 4 contains details of estimation methods for our efficacy measures, followed by conclusions and suggestions for further work in Section 5.

2. MODELS FOR OUTBREAKS IN PAIRS AND VACCINE EFFECTS

Consider the person-to-person transmission of infection in a community of households, over a specific observation period. A study consists of observing outbreaks in households with two susceptible individuals, where one or both may be vaccinated. It is assumed that the study period is long, relative to the duration of a household outbreak, and that the size of a household outbreak is observed even when its primary case occurs towards the end of the observation period. Larger households may be included provided other household members are fully immune, typically as a result of previous exposure to the infection.

2.1. Outbreaks in unvaccinated pairs

We begin with assumptions similar to those made in Reference [12]. Consider first a household pair in which neither partner is vaccinated. Every individual is exposed to a global force of infection over the study period and may be exposed to a force of infection from an infected partner. Each individual avoids infection from *global* contacts, independently, with probability q_g. The probability of avoiding infection from an infected *household* partner is q_h. In each household pair the possible number infected during the observation period is 0, 1 or 2, with probabilities

$$p_0(0) = q_g^2, \quad p_0(1) = 2q_g(1-q_g)q_h \quad \text{and} \quad p_0(2) = 2q_g(1-q_g)(1-q_h) + (1-q_g)^2$$

respectively, where the subscript in p_0 indicates that neither individual in the pair is vaccinated.

2.2. Vaccine response and efficacy measures

We now recall a framework for vaccine response described by Becker and Starczak [13]. For an individual who is vaccinated we describe the vaccine response by a realization of the random vector (A,B). The random variable A describes the relative susceptibility compared to an unvaccinated individual, and B the relative infectivity, should the vaccinee become infected. Vaccine responses are independently and identically distributed for different individuals, but A and B of the same individual may be correlated. To make this more precise, consider a vaccinated individual with realized vaccine response (a,b) and an unvaccinated household partner. As a result of vaccination, the force of infection acting on that individual at time t is changed from λ_t to $a\lambda_t$. It follows that the probability of avoiding infection from outside the household becomes q_g^a and the probability of avoiding infection from an infected unvaccinated household partner is changed to q_h^a. If this vaccinated individual does become infected, from a global source, then this vaccinee infects the unvaccinated partner with probability $1-q_h^b$. Unconditionally, these probabilities are $E(q_g^A)$, $E(q_h^A)$ and $E(1-q_h^B)$, respectively.

The full vaccine response is described by the probability distribution of (A,B). Three summary measures of the vaccine response, with regard to susceptibility and infectivity, are of particular interest. One is the average reduction in susceptibility, per contact, given by

$$\mathrm{VE_S} = 1 - E(A)$$

and called the protective vaccine efficacy. Note that $\mathrm{VE_S} = 0$ corresponds to no protective effect and $\mathrm{VE_S} = 1$ to complete protection. Measures of a vaccine's effect on infectivity and transmissibility requires more care because infectivity is only relevant if the individual gets infected. For example, the value of B is irrelevant in the extreme case when $A = 0$, since then the individual is never infected. Similarly, the value of B has a minor impact when A is positive, but very small, because the individual is rarely infected. On the other hand, when the value of A is large, near 1, then a small value for B can reduce community transmission substantially. A measure which can reflect this is $E(A) - E(AB)$. Rescaling, so that 0 means no reduction in infectivity and 1 corresponds to no infectivity remains, gives

$$\mathrm{VE_I} = 1 - E(AB)/E(A)$$

The fact that VE_I is not defined when $E(A)=0$ is not of concern, because this generally means that $\Pr(A=0)=1$, so that no vaccinated individuals become infected. The third summary measure of interest is

$$VE_{SI}=1-E(AB)$$

which measures the dual effect of changes in susceptibility and infectivity on transmission of infection. Note that

$$1-VE_{SI}=(1-VE_S)(1-VE_I)$$

so that estimation of two of these summary measures determines the third.

A desirable property of these measures of vaccine efficacy, in contrast to measures based on relative risk, is that an estimate obtained in the setting considered here can be meaningfully interpreted for use of the vaccine in a different settings. In other words, they are primarily characteristics of the vaccine and do not depend on characteristics, such as community setting and duration of study period, which are specific to the conducted study.

Some of the discussion below refers to the specific vaccine response model described by

$$\Pr(A=0)=c,\quad \Pr(A=a,B=b)=d\quad \text{and}\quad \Pr(A=1,B=1)=f \tag{1}$$

where $0<a,b<1$ and $c+d+f=1$. This *all-partial-none* vaccine response model was previously used in Reference [6]. Its parameters capture important features of a vaccine response, namely the proportion of individuals that become completely immune (c), the proportion of vaccinations that fail (f) and, for individuals who become partially protected, it captures the vaccine effects on susceptibility (a) and infectivity (b).

Two particular responses contained in model (1) are mentioned below, namely the *partial* response model obtained when $d=1$ and the *all-none* response model obtained when $d=0$. In the *partial* response all individuals have the same response (a,b). It is sometimes referred to as a 'leaky' response; see Reference [14].

2.3. Outbreaks in vaccinated pairs

In household pairs with both individuals vaccinated prior to the observation period the number infected can be 0, 1 or 2, with probabilities

$$p_2(0)=[E(q_g^A)]^2$$

$$p_2(1)=2E[(1-q_g^{A_1})q_g^{A_2}q_h^{B_1A_2}]\quad \text{and}$$

$$p_2(2)=2E[(1-q_g^{A_1})q_g^{A_2}(1-q_h^{B_1A_2})]+[E(1-q_g^A)]^2$$

respectively, where (A_1,B_1) and (A_2,B_2) are assumed to be independent random vectors. The possibility of dependent vaccine responses within households, perhaps because pair members are genetically similar, is not allowed for.

There are four outcomes for a pair in which one individual is vaccinated and the other is not, namely neither is infected, only the unvaccinated individual is infected, only the vaccinated

individual is infected or both are infected. The probabilities of these events, respectively, are

$$p_1(0,0) = q_g E(q_g^A)$$

$$p_1(1,0) = (1-q_g)E(q_g^A q_h^A)$$

$$p_1(0,1) = q_g E[(1-q_g^A)q_h^B] \quad \text{and}$$

$$p_1(1,1) = (1-q_g)E(1-q_g^A q_h^A) + q_g E[(1-q_g^A)(1-q_h^B)]$$

where the subscript in p_1 indicates that one individual of the pair is vaccinated.

Koopman and Little [7], Longini *et al.* [8] and Datta *et al.* [10, 11] also study partially protective vaccine effects, but in their models vaccination acts linearly on infection probabilities. The present vaccine response model affects the avoidance probability by a factor in the exponent. In other words, vaccination acts linearly on the logarithm of the avoidance probability. It has the advantage that the interpretations of vaccine effects described by this model do not depend on the duration of the study.

3. EFFICACY MEASURES BASED ON RELATIVE ATTACK RATES

Before considering estimation of the above measures of vaccine efficacy we present examples, to show that estimates of the form

$$\widehat{\text{VE}} = 1 - \text{relative attack rate} = 1 - \frac{\text{proportion of vaccinated individuals infected}}{\text{proportion of unvaccinated individuals infected}}$$

can seriously mislead us about vaccine efficacy. The examples serve two related purposes. Firstly, they provide an important alert to the many researchers who use and promote this type of estimate; see for example References [2, 5, 7, 15]. The reason for the popularity of $\widehat{\text{VE}}$ lies in its easy computation and its apparent simple interpretation, although the latter is a misguided impression. The second purpose of these examples is to provide a compelling argument for the need to develop alternative methods of estimation, and associated vaccine trial designs, such as those proposed in this paper. Some guidance on settings for which $\widehat{\text{VE}}$ is appropriate, and alternative measures when it is not, is given by Hernández and Castillo [16] and Hernández [17].

Smith *et al.* [15] point out that the estimate $\widehat{\text{VE}}$ depends on the type of vaccine response, which in itself is a concern. In particular, they point out that the cumulative force of infection acting over the observation period has a substantial effect on the value obtained for $\widehat{\text{VE}}$ when the vaccine response includes a partial component, e.g. when $d > 0$ in equation (1). The discussion by Smith *et al.* does not focus on transmission, so it is worth pointing out that for infectious diseases the cumulative force of infection depends substantially on transmission characteristics of the disease and the community setting. It follows that the value of $\widehat{\text{VE}}$ depends substantially on transmission characteristics of the disease and the community setting. The following two examples highlight this.

Example 1 (*Outbreaks in household pairs*)
For a large number of household pairs, each with one vaccinated and one unvaccinated individual, we can deduce that $\widehat{\mathrm{VE}}$ estimates

$$\mathrm{VE} = 1 - \frac{p_1(0,1) + p_1(1,1)}{p_1(1,0) + p_1(1,1)}$$

where the probabilities $p_1(i,j)$ are as in Section 2.3.

For the all-none vaccine response $\Pr(A=1,B=1) = f = 1 - \Pr(A=0)$ and a sensible measure of protective vaccine efficacy is $1-f$. However,

$$\mathrm{VE} = \frac{1-f}{1+q_{\mathrm{g}}(1-q_{\mathrm{h}})f}$$

and as $\mathrm{VE} \leqslant 1-f$, we see that $\widehat{\mathrm{VE}}$ generally underestimates the desired quantity $1-f$ and can estimate a quantity as low as one half of this value, when $q_{\mathrm{g}} \approx 1$, $q_{\mathrm{h}} \approx 0$ and $f \approx 1$. In other words, $\widehat{\mathrm{VE}}$ may seriously mislead us even for an all-none vaccine response when the community consists of households *and* households are recruited into the study on the basis of the vaccination status of its members. Note, however, that results in Hernández-Suárez [17] imply that $\widehat{\mathrm{VE}}$ gives a satisfactory estimate of vaccine efficacy for the all-none vaccine response when the vaccination status is determined independently for each member of the recruited households.

For the partial vaccine response, described by $\Pr(A=a,B=b)=1$, we find

$$\mathrm{VE} = 1 - \frac{q_{\mathrm{g}}^a q_{\mathrm{h}}^a (1-q_{\mathrm{g}}) - q_{\mathrm{g}} q_{\mathrm{h}}^b (1-q_{\mathrm{g}}^a)}{1 - q_{\mathrm{g}}^{a+1} - q_{\mathrm{g}} q_{\mathrm{h}}^b (1-q_{\mathrm{g}}^a)}$$

As an illustration let $q_{\mathrm{g}}=0.9$, $q_{\mathrm{h}}=0.1$ and $a=0.9$, so that the probability of being infected from a global source is small, the within-household transmission rate is high when there is an infective in the household and the vaccine offers a modest level of protection. If b is sufficiently small, corresponding to substantial infectivity reduction for the vaccinee, it follows that the smaller global force of infection exerted on vaccinated individuals (due to the protective effect of the vaccine) is more than compensated for by the possible high force of infection from an infected unvaccinated partner. In other words, the average cumulative force of infection exerted on unvaccinated individuals is smaller than that exerted on vaccinated individuals. Indeed, calculations show that VE increases with b, from -0.70 when $b=0$ (vaccination removes all infectivity) to 0.019 when $b=1$ (vaccination does not affect infectivity). Therefore, a vaccine that reduces susceptibility and infectivity can expect to return a value of $\widehat{\mathrm{VE}}$ that is negative, suggesting that vaccination makes individuals more vulnerable to infection when it clearly does not. This illustrates again that $\widehat{\mathrm{VE}}$ can seriously mislead us, and it occurs because $\widehat{\mathrm{VE}}$ does not take transmission into account.

The next example presents a different kind of concern.

Example 2 (*A sample of affected households*)
A change in infectivity is difficult to estimate, because the source of an infection is generally unknown. To overcome this dilemma, investigators sometimes sample infected households in

which the vaccination status of primary cases can be established. Then an estimate of the form $\widehat{\mathrm{VE}}$ is applied to data on first-generation secondary cases. Specifically,

$$\widehat{\mathrm{VE}}' = 1 - p_{\mathrm{v}}/p_{\mathrm{u}}$$

is used to estimate the reduction in infectivity, where p_{v} is the observed proportion of cases among unvaccinated household members exposed to a vaccinated primary household case and p_{u} is the observed proportion of cases among unvaccinated household members exposed to an unvaccinated primary household case. Halloran *et al.* [5] analyse data from a study of this type.

A serious concern arises with such studies when the vaccine response may give partial immunity, as for the all-partial-none response models with $d>0$, as is now explained. Partial immunity generally implies some reduction in both susceptibility and infectivity. In other words, a vaccinated individual has A and B values in $(0,1)$, and A and B are typically positively correlated. The problem is that vaccinated primary cases of the sampled households are not representative of vaccinees. They are infected vaccinees, so they are likely to include a larger proportion with high susceptibility and infectivity than a group of randomly selected vaccinees.

To show that this kind of study can give a misleading assessment of the effect of vaccination on infectivity, consider a household pair with one member vaccinated, who had the specific vaccine response $(A,B)=(a,b)$. Being interested in primary cases, we consider global forces of infection λ_t and $a\lambda_t$ acting on the unvaccinated and vaccinated individuals, respectively, at time t during the recruitment period. Given that both members of a household pair remain uninfected at time t and one in the pair is infected during the time increment $(t,t+\mathrm{d}t)$, the probability that it is the vaccinated individual is $a\lambda_t\,\mathrm{d}t/(a\lambda_t\,\mathrm{d}t+\lambda_t\,\mathrm{d}t)=a/(a+1)$.

Now assume a partial-none response to vaccination described by

$$\Pr(A=1,B=1)=1-\Pr(A=a,B=b)=f \quad \text{where } 0<a,b<1$$

Then, among all household pairs with one vaccinee, a fraction f have a vaccinee at the higher susceptible level. However, among household pairs with one vaccinee who is a primary household case, the fraction

$$\frac{\frac{1}{2}f}{\frac{1}{2}f+[a/(a+1)](1-f)}=\frac{f+af}{f+2a-af}$$

has a vaccinee at the higher susceptible level. To illustrate suppose that the probability of a vaccine failure is $f=0.2$ and $a=0.1$, which means that successful vaccination reduces the chance of infection during a contact with an infective to one-tenth of the chance for an unvaccinated individual. Overall, 20 per cent of vaccinated individuals have a vaccine failure, whereas 57.9 per cent of vaccinated primary cases had a vaccine failure. Estimating $\mathrm{VE_I}$ by treating vaccinated primary cases as typical vaccinees is clearly going to severely underestimate the reduction in infectivity induced by vaccination.

These examples clearly demonstrate that a fair assessment of vaccine effects requires a method of estimation that acknowledges that outbreak data are the result of disease transmission between individuals.

4. ESTIMATING VACCINE EFFECTS FROM HOUSEHOLD PAIRS DATA

Suppose the study has n_j household pairs with j individuals vaccinated, ($j=0,1,2$). Let $n_0(i)$ denote the number of households having no vaccinated individuals in which i individuals became infected ($i=0,1,2$), and similarly define $n_2(i)$ for households with two vaccinated individuals. Also, let $n_1(i,j)$ denote the number of households with one vaccinated individual in which the outbreak consists of i unvaccinated cases and j vaccinated cases ($i,j=0,1$). Treating households as independent, the log-likelihood for these data is

$$\ell(q_g, q_h, \boldsymbol{\theta}) = \sum_{i=0}^{2} n_0(i) \ln[p_0(i)] + \sum_{i=0}^{1}\sum_{j=0}^{1} n_1(i,j) \ln[p_1(i,j)] + \sum_{i=0}^{2} n_2(i) \ln[p_2(i)] \tag{2}$$

where $\boldsymbol{\theta}$ is the vector of parameters of the vaccine response distribution. This is the log-likelihood function for data from three multinomial distributions, and we can clearly estimate $\{p_0(i)\}$, $\{p_1(i,j)\}$ and $\{p_2(i)\}$. Our interest, however, lies in expressing these cell probabilities in terms of parameters that describe transmission and vaccine effects, namely $(q_g, q_h, \boldsymbol{\theta})$, and making inferences about certain characteristics of the distribution of (A,B). We now present some inference results that do not require us to specify a form for the random vaccine response distribution.

4.1. Distribution-free inferences about vaccine effects

The outcome probabilities, given in Sections 2.1 and 2.3, and the log-likelihood function (2) indicate that outbreak data from household pairs with 0, 1 and 2 vaccinated members enable the estimation of q_g, q_h and

$$\boldsymbol{\theta} = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \end{pmatrix} = \begin{pmatrix} E(q_g^A) \\ E(q_g^A q_h^A) \\ E[(1-q_g^A) q_h^B] \\ E[(1-q_g^{A_1}) q_g^{A_2} q_h^{B_1 A_2}] \end{pmatrix}$$

as well as functions of these. The dimension of $(q_g, q_h, \boldsymbol{\theta})$ is one less than the number of degrees of freedom for these three multinomial distributions, because $p_2(0) = [p_1(0,0)]^2 / p_0(0)$ holds without any assumptions about the vaccine response distribution.

Maximum likelihood estimates can be obtained by substituting

$$p_0(0) = q_g^2, \quad p_0(1) = 2q_g(1-q_g)q_h, \quad p_0(2) = 2q_g(1-q_g)(1-q_h) + (1-q_g)^2$$

$$p_1(0,0) = q_g\theta_1, \quad p_1(1,0) = (1-q_g)\theta_2, \quad p_1(0,1) = q_g\theta_3, \quad p_1(1,1) = 1-\theta_2-q_g(\theta_1-\theta_2+\theta_3)$$

$$p_2(0) = \theta_1^2, \quad p_2(1) = 2\theta_4 \quad \text{and} \quad p_2(2) = 1 - \theta_1^2 - 2\theta_4$$

into the log-likelihood (2) and maximizing with respect to $(q_g, q_h, \boldsymbol{\theta})$. Under the assumption $A \leqslant 1$ and $B \leqslant 1$, so that vaccination does not *increase* susceptibility or infectivity, the constraints $\theta_1 \geqslant \theta_2$ and $\theta_3 \geqslant \theta_4$ need to be imposed when maximizing the log-likelihood function.

Consistent estimates of VE_S, VE_I and VE_{SI} are generally not available, because these measures cannot be expressed as functions of $(q_g, q_h, \boldsymbol{\theta})$ without making the vaccine response model more specific. The outbreak data are nevertheless informative about VE_S, VE_I and VE_{SI}, because each of them can be bounded by estimable quantities. Below we give bounds that can be estimated by substituting maximum likelihood estimates for $(q_g, q_h, \boldsymbol{\theta})$.

4.1.1. Reduction in susceptibility. Using convexity arguments as in Reference [18] and the assumptions $A \leqslant 1$ and $B \leqslant 1$, it can be shown that

$$\frac{\log[E(q_g^A)]}{\log(q_g)} \leqslant E(A) \leqslant \frac{1 - E(q_g^A)}{1 - q_g}$$

Translating these bounds to $VE_S = 1 - E(A)$ gives the estimable bounds

$$\frac{\theta_1 - q_g}{1 - q_g} \leqslant VE_S \leqslant \frac{\log(q_g/\theta_1)}{\log(q_g)} \tag{3}$$

The bounds are sharp, with the lower bound attained for the all-none response and the upper bound attained for the partial response. The two bounds are extremes in the sense that, for a given $E(A)$, the upper bound is attained when $\mathrm{Var}(A) = 0$ while the lower bound is attained when $\mathrm{Var}(A)$ has its maximum value of $E(A)[1 - E(A)]$. It is interesting that corresponding bounds for VE_S based on infection data from a uniformly-mixing community of homogeneous individuals, see Becker and Utev [18], are attained by precisely the same vaccine responses.

There are two important applications of the inequalities (3). One is to obtain an interval estimate for VE_S that does not depend upon a particular vaccine response model for (A, B). For this we estimate each of the bounds. The second is robust testing of the hypothesis that the vaccine has a protective effect, by testing H_0: $\varphi = 0$ against H_0: $\varphi > 0$, where $\varphi = (\theta_1 - q_g)/(1 - q_g)$. When the number of pairs is large, this test can be conducted using the large sample normality of the maximum likelihood estimate $\hat{\varphi}$, with the standard deviation given below.

To illustrate that the bounds in (3) provide estimates of practical value we compute, in Table I, the lower and upper bounds of VE_S for some plausible parameter values under the protective vaccine response described by $\Pr(A = 1) = f = 1 - \Pr(A = a)$. The small difference between these bounds indicates that estimation of these bounds is informative about VE_S. Indeed the difference is likely to be substantially smaller than the width of the confidence interval for the bound.

We now argue that the bounds on VE_S have practical value even when no specific protective vaccine effect is assumed. Denote the difference between the upper and lower bounds by $f(q, \theta)$ when $q_g = q$ and $\theta_1 = \theta$. Then

$$f(q, \theta) = \frac{1 - \theta}{1 - q} - \frac{\log \theta}{\log q}$$

Suppose that q_g is not smaller than some fixed value q_L. For $q_L \leqslant q \leqslant \theta \leqslant 1$, f is maximized at $q_{\max} = q_L$ and $\theta_{\max} = (q_L - 1)/\log q_L$. Figure 1 illustrates $f(q_{\max}, \theta_{\max})$, i.e. the maximum possible difference between the bounds, for a range of values of q_L. Note that this difference is small for all values of q_L greater than about 0.2. The bounds are therefore of practical

Table I. Bounds for VE_S given by (3), for partial-none vaccine protection and $q_g = 0.9$.

		a		
		0.1	0.3	0.5
f	0.05	(0.850, 0.857)	(0.654, 0.666)	(0.462, 0.476)
	0.15	(0.761, 0.770)	(0.586, 0.598)	(0.414, 0.427)
	0.25	(0.671, 0.683)	(0.517, 0.530)	(0.365, 0.377)

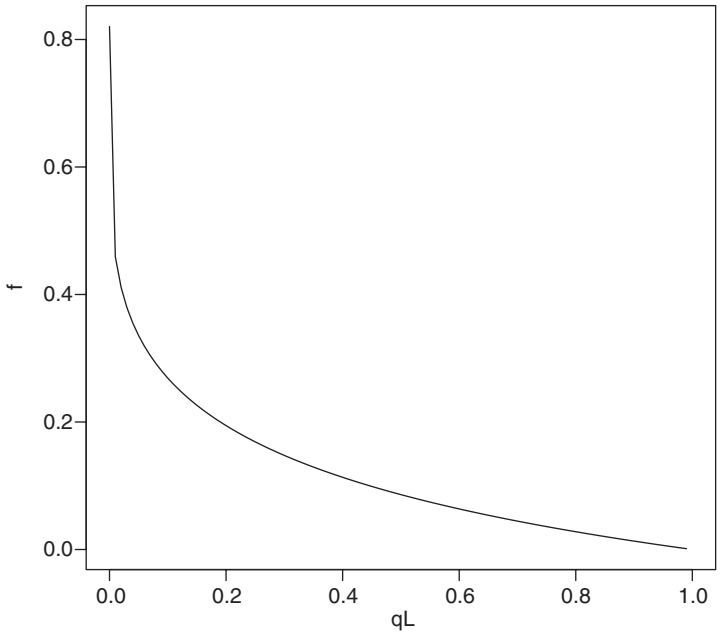

Figure 1. Maximum possible difference, f, between upper and lower bounds for VE_S, as q_L varies. Note that $f \uparrow 1$ as $q_L \downarrow 0$.

value whenever less than 80 per cent of susceptible individuals become infected during the observation period.

An approximate assessment of the precision of maximum likelihood estimates of the bounds is possible by noting that the bounds depend only on q_g and θ_1. Inspection of the expressions for the $p_0(i)$, the $p_1(i,j)$ and the $p_2(i)$ reveals that the information about q_g and θ_1 is essentially contained in the cell frequencies $n_0(0)$, $n_1(0,0)$ and $n_2(0)$. They are observations on independent Binomial(n_0, q_g^2), Binomial$(n_1, q_g\theta_1)$ and Binomial(n_2, θ_1^2) distributions, respectively. Suppose that our data consists only of observing $n_0(0)$, $n_1(0,0)$ and $n_2(0)$ uninfected households out of n_0, n_1 and n_2, respectively. Let $n = n_0 + n_1 + n_2$ and for $i = 0, 1, 2$ let $\pi_i := \lim_{n\to\infty} n_i/n > 0$, the large-sample proportion of pairs with i vaccinated partners. Then

the large-sample variance–covariance matrix is

$$\Sigma^2(\hat{q}_g, \hat{\theta}_1) \approx \frac{1}{\Delta n} \begin{pmatrix} \dfrac{4\pi_2}{1-\theta_1^2} + \dfrac{\pi_1 q_g}{\theta_1(1-q_g\theta_1)} & \dfrac{-\pi_1}{1-q_g\theta_1} \\ \dfrac{-\pi_1}{1-q_g\theta_1} & \dfrac{4\pi_0}{1-q_g^2} + \dfrac{\pi_1\theta_1}{q_g(1-q_g\theta_1)} \end{pmatrix}$$

where

$$\Delta = \frac{16\pi_0\pi_2}{(1-q_g^2)(1-\theta_1^2)} + \frac{4\pi_1\pi_2\theta_1}{q_g(1-q_g\theta_1)(1-\theta_1^2)} + \frac{4\pi_0\pi_1 q_g}{\theta_1(1-q_g\theta_1)(1-q_g^2)}$$

The delta method now gives the large-sample standard deviations

$$\text{s.d.}(\hat{\varphi}) = \text{SD}\left(\frac{\hat{\theta}_1 - \hat{q}_g}{1-\hat{q}_g}\right) \approx \left[\frac{(\theta_1-1)^2}{(1-q_g)^4}\Sigma_{11}^2 + 2\frac{(\theta_1-1)}{(1-q_g)^3}\Sigma_{12}^2 + \frac{1}{(1-q_g)^2}\Sigma_{22}^2\right]^{1/2} \quad \text{and}$$

$$\text{SD}\left(\frac{\log(\hat{q}_g/\hat{\theta}_1)}{\log(\hat{q}_g)}\right) \approx \left[\frac{(\log\theta_1)^2}{q_g^2(\log q_g)^4}\Sigma_{11}^2 - 2\frac{\log\theta_1}{q_g\theta_1(\log q_g)^3}\Sigma_{12}^2 + \frac{1}{\theta_1^2(\log q_g)^2}\Sigma_{22}^2\right]^{1/2}$$

for the estimated lower and upper bounds for VE_S, respectively.

From these standard deviations it can be shown that studies involving a few hundred pairs of individuals can be large enough to give estimates of reasonable precision. For example, with $n_0 = n_1 = n_2 = 100$, $\theta_1 = 0.9$ and setting $q_g = 0.8, 0.7$ and 0.6, the standard deviation of the lower bound is 0.103, 0.064 and 0.045, respectively, and for the upper bound we obtain 0.106, 0.062 and 0.041. This suggests that our estimation method is useful for data obtained from studies of a size that are feasible in practice.

4.1.2. Reduction in infectivity. Bounds for $E(AB)$ can be obtained as follows. Fix $0 < q < 1$ and $0 \leqslant z \leqslant 1$. Then for $0 \leqslant x \leqslant 1$, simple geometry shows that

$$(q^z \log q)x + q^z(1 - z\log g) \leqslant q^x \leqslant -(1-q)x + 1$$

Applying these inequalities with x replaced by A or B, and q replaced by q_g or q_h, respectively, yields bounds for A and B. Multiplying these bounds and taking expectations (with $z = 0$, for example), gives

$$\frac{E[(1-q_g^A)(1-q_h^B)]}{\log q_g \log q_h} \leqslant E(AB) \leqslant \frac{E[(1-q_g^A)(1-q_h^B)]}{(1-q_g)(1-q_h)} \tag{4}$$

In principle, other bounds can be obtained by taking a different value for z, although calculations are then complicated by the fact that the bounds can become negative, and hence redundant. Note that since $E[(1-q_g^A)(1-q_h^B)] = 1 - \theta_1 - \theta_3$, the bounds in (4) are estimable quantities. The upper bound is attained by the all-none vaccine response, but the lower bound is not attained for any explicit response model. The two bounds coincide as $q_g, q_h \to 1$, and are reasonably close to each other when q_g and q_h are not too small.

Table II. Bounds for VE_{SI} and VE_I given by (5) and (6), for partial-none vaccine response with $q_g = 0.9$, $q_h = 0.7$ and $f = 0.15$.

		a = 0.1	a = 0.3	a = 0.5
VE_{SI}				
	0.1	(0.840, 0.872)	(0.819, 0.856)	(0.799, 0.840)
b	0.3	(0.820, 0.856)	(0.761, 0.809)	(0.702, 0.762)
	0.5	(0.801, 0.841)	(0.706, 0.765)	(0.613, 0.691)
VE_I				
	0.1	(0.302, 0.464)	(0.550, 0.652)	(0.650, 0.726)
b	0.3	(0.216, 0.399)	(0.404, 0.539)	(0.481, 0.595)
	0.5	(0.136, 0.337)	(0.268, 0.434)	(0.324, 0.472)

Translating these bounds to $VE_{SI} = 1 - E(AB)$ gives

$$1 - \frac{1-\theta_1-\theta_3}{(1-q_g)(1-q_h)} \leqslant VE_{SI} \leqslant 1 - \frac{1-\theta_1-\theta_3}{\log q_g \log q_h} \tag{5}$$

Finally, combining the bounds for $E(A)$ and $E(AB)$ we find

$$1 - \frac{(1-\theta_1-\theta_3)\log q_g}{(1-q_g)(1-q_h)\log \theta_1} \leqslant VE_I \leqslant 1 - \frac{(1-q_g)(1-\theta_1-\theta_3)}{(1-\theta_1)\log q_g \log q_h} \tag{6}$$

The bounds for $VE_I = 1 - E(AB)/E(A)$ are not sharp because they consist of combinations of two bounds that are attained for different vaccine responses.

Table II gives upper and lower bounds for VE_I and VE_{SI}, for some typical parameter values. In most cases the bounds are tight enough to be of practical value, especially so for VE_{SI}.

4.2. *Inferences for all-partial-none vaccine responses*

The choice of parameter $\boldsymbol{\theta}$ was made primarily for reasons of convenience after inspecting expressions for the cell probabilities of our transmission models. It is worthwhile considering a change of parameter that directly reflects vaccine response characteristics of interest. This is achieved by adopting the all-partial-none vaccine response model given by (1). Adopting this model corresponds to a re-parameterization from $(q_g, q_h, \boldsymbol{\theta})$ to (q_g, q_h, a, b, c, f), so the dimension of the parameter has not changed.

Maximum likelihood estimates of (q_g, q_h, a, b, c, f) can, in principle, be obtained by writing the $p_0(i)$, $p_1(i,j)$ and $p_2(i)$ in terms of (q_g, q_h, a, b, c, f) and maximizing the log-likelihood for the observed data with respect to these parameters. Confidence intervals can be obtained by bootstrap methods. In practice such calculations are not entirely trivial, due to the large number of parameters and the relatively complicated way these probabilities depend on (q_g, q_h, a, b, c, f). Maximum likelihood estimation becomes a little easier when we are prepared to make more specific assumptions about the vaccine response. For example, with an all-none vaccine response, given by $d = 0$, it can be shown that

$$VE_S = 1 - \frac{1-\theta_1}{1-q_g}$$

$$\mathrm{VE_I} = 1 - \frac{1-\theta_1-\theta_3}{(1-\theta_1)(1-q_\mathrm{h})} \quad \text{and}$$

$$\mathrm{VE_{SI}} = 1 - \frac{1-\theta_1-\theta_3}{(1-q_\mathrm{g})(1-q_\mathrm{h})}$$

This makes maximum likelihood estimation of these measures a little easier, because the $p_0(i)$, $p_1(i,j)$ and $p_2(i)$ have relatively simple expressions in terms of $(q_\mathrm{g}, q_\mathrm{h}, \boldsymbol{\theta})$. The same approach works for the partial vaccine response, given by setting $d=1$, because then the vaccine measures are given by

$$\mathrm{VE_S} = 1 - \frac{\log\theta_1}{\log q_\mathrm{g}}$$

$$\mathrm{VE_I} = 1 - \frac{\log\theta_3 - \log(1-\theta_1)}{\log q_\mathrm{h}} \quad \text{and}$$

$$\mathrm{VE_{SI}} = 1 - \frac{\log\theta_1[\log\theta_3 - \log(1-\theta_1)]}{\log q_\mathrm{g}\log q_\mathrm{h}}$$

For these more specific vaccine response models one can also write down some explicit expressions for consistent, although not fully efficient, estimates of interest. Suppose we are prepared to adopt the all-none response. Then a consistent estimate of $E(A)=f$ is given by

$$\hat{f} = \frac{[n_1(0,1)+n_1(1,1)]/n_1}{[n_0(1)+2n_0(2)]/n_0}$$

Suppose, instead, that we are prepared to assume that the partial response applies. Then one consistent estimate of $E(A)=a$ is given by

$$\hat{a} = \frac{\log[n_2(0)/n_2]}{\log[n_0(0)/n_0]}$$

although this is easily improved upon by adding the information that $n_1(0,0)/n_1$ contains about a.

In practice it seems best to retain the all-partial-none vaccine response model (1), for which we recommend either using the distribution-free bounds of the previous subsection (they apply to an arbitrary vaccine response) or making Bayesian inferences implemented by Markov chain Monte Carlo methods.

Estimation methods for the all-partial-none response model in a Bayesian framework, using MCMC methods, are described in detail in Reference [6]. One finding of particular relevance here is that both $E(A)$ and $E(AB)$ (and hence $\mathrm{VE_S}$ and $\mathrm{VE_{SI}}$) are usually estimable with precision, meaning that their posterior densities typically have small standard deviation. The methods also allow exploration of the posterior density of the ratio $E(AB)/E(A)$, and this is also found to be estimable with reasonable precision. For example, setting $a=0.4, b=0.5, c=0.45$, $f=0.15, q_\mathrm{g}=0.8$, $q_\mathrm{h}=0.7$, $n_0=n_1=n_2=100$ and using data consisting of the expected frequency for each outcome, we found parameter posterior modes in close agreement with the known true values. Moreover, posterior modes for $E(A)=0.31$, $E(AB)=0.23$ and

$E(AB)/E(A) = 0.74$ were, respectively, $0.32, 0.26$ and 0.82, with corresponding standard deviations $0.074, 0.089$ and 0.19.

5. DISCUSSION

In contrast to measures based on relative attack rates, the measures VE_S, VE_I and VE_{SI} considered here indicate the same efficacy irrespective of the community setting and the duration of the study period. This is achieved by defining them solely in terms of characteristics of the distribution of vaccine responses, and viewing infectious disease data with reference to a model that captures transmission together with the effect that vaccine responses have on transmission. A challenge associated with these efficacy measures is that an estimation procedure needs to be developed for each specific study setting.

Here we have considered estimation of these vaccine efficacy measures from a study based on outbreaks in household pairs. This study design provides two advantages. Firstly, it enables estimation of the vaccine effect on infectivity, because such outbreaks contain some information about the possible source of infection. Secondly, transmission models for pairs are not encumbered by the complexity present in outbreaks in larger households.

Becker and Utev [18] provide a way of estimating VE_S by use of estimable bounds for a study in which every individual is exposed to the same force of infections throughout the study period. This approach estimates VE_S in a way that is robust with respect to the nature of the vaccine response. Here we provide similar estimable bounds for VE_S corresponding to a household pairs study design, and extend the approach to estimation of VE_I and VE_{SI} for that setting.

A further potential attraction of the household pairs study design is that individuals of the same pair are matched with regard to geographic location. This can be important when there are many pairs in the study, because the global force of infection may differ appreciably between locations of residence. It suggests the need to permit heterogeneity in the global infection probability q_g. A natural way to incorporate heterogeneity is to allow each pair's q_g value to be a realization from some probability distribution, whose parameters we then seek to estimate. For certain choices of distribution it turns out that an MCMC approach to estimation can be applied in a convenient way, and moreover this approach is also feasible for data on households of arbitrary size. The details of this will be presented elsewhere. Similar comments apply to heterogeneity in q_h, which may arise from susceptibility or infectivity varying among individuals or between households.

Another area for further work is that of model choice. In the present paper, we describe inference methods for both distribution-free and parametric vaccine response models. If it is desired to use a parametric model, which is the most appropriate? For example, is there an appreciable gain, when estimating VE_S, VE_I and VE_{SI}, by using the all-partial-none vaccine response model rather than the partial-none model ($c = 0$)? There are various ways to answer this question, and these will be addressed elsewhere.

REFERENCES

1. Clements J, Brenner R, Rao M, Tafari N, Lowe C. Evaluating new vaccines for developing countries: efficacy or effectiveness? *Journal of the American Medical Association* 1996; **275**:391–397.

2. Orenstein WA, Bernier RH, Dondero TJ *et al*. Field evaluation of vaccine efficacy. *Bulletin of the World Health Organization* 1985; **63**:1055–1068.
3. Chen RT, Orenstein WA. Epidemiologic methods in immunization programs. *Epidemiologic Reviews* 1996; **18**:99–117.
4. Halloran ME, Struchiner CJ, Longini IM. Study designs for evaluating different efficacy and effectiveness aspects of vaccines. *American Journal of Epidemiology* 1997; **146**:789–803.
5. Halloran ME, Préziosi M-P, Chu H. Estimating vaccine efficacy from secondary attack rates. *Journal of the American Statistical Assocociation* 2003; **98**:38–46.
6. Becker NG, Britton T, O'Neill PD. Estimating vaccine effects on transmission of infection from household data. *Biometrics* 2003; **59**:467–475.
7. Koopman JS, Little RJ. Assessing HIV vaccine effects. *American Journal of Epidemiology* 1995; **142**:1113–1120.
8. Longini IM, Datta S, Halloran ME. Measuring vaccine efficacy for both susceptibility to infection and reduction in infectiousness for prophylactic HIV-1 vaccines. *Journal of Acquired Immune Deficiency Syndrome* 1996; **13**:440–447.
9. Rida W. Assessing the effect of HIV vaccination on infectiousness. *Statistics in Medicine* 1996; **15**:2393–2404.
10. Datta S, Halloran ME, Longini IM. Augmented HIV vaccine trial design for estimating reduction in infectiousness and protective efficacy. *Statistics in Medicine* 1998; **17**:185–200.
11. Datta S, Halloran ME, Longini IM. Efficiency of estimating vaccine efficacy for susceptibility and infectiousness: randomization by individual *versus* household. *Biometrics* 1999; **55**:792–798.
12. Longini IM, Koopman JS. Household and community transmission parameters from final distributions of infections in households. *Biometrics* 1982; **38**:115–126.
13. Becker NG, Starczak DN. The effect of random vaccine response to prevent epidemics. *Mathematical Biosciences* 1998; **154**:117–135.
14. Halloran ME, Haber M, Longini IM. Interpretation and estimation of vaccine efficacy under heterogeneity. *American Journal of Epidemiology* 1992; **136**:328–343.
15. Smith PG, Rodrigues LC, Fine PEM. Assessment of the protective efficacy of vaccines against common diseases using case-control and cohort studies. *International Journal of Epidemiology* 1984; **13**:87–96.
16. Hernández-Suárez CM, Castillo-Chavez C. Urn models and vaccine efficacy estimation. *Statistics in Medicine* 2000; **19**:827–835.
17. Hernández-Suárez CM. A note on the distribution of the number of vaccinated infected under non-random mixing. *Statistics in Medicine* 2001; **20**:1983–1986.
18. Becker NG, Utev S. Protective vaccine efficacy when vaccine response is random. *Biometrical Journal* 2002; **44**:29–42.
19. Andersson H, Britton T. *Stochastic Models and Their Statistical Analysis*, Springer Lecture Notes in Statistics, vol. 151. Springer: New York, 2000.

