Estimating Vaccine Effects on Transmission of Infection from Household Outbreak Data

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, University of Nottingham, Nottingham NG7 2RD, U.K.
*email: pdo@maths.nott.ac.uk

SUMMARY. This article is concerned with a method for making inferences about various measures of vaccine efficacy. These measures describe reductions in susceptibility and in the potential to transmit infection. The method uses data on household outbreaks; it is based on a model that allows for transmission of infection both from within a household and from the outside. The use of household data is motivated by the hope that these are informative about vaccine-induced reduction of the potential to transmit infection, as household outbreaks contain some information about the possible source of infection. For illustration, the method is applied to observed data on household outbreaks of smallpox. These data are of the form needed and the number of households is of a size that can be managed in a vaccine trial. It is found that vaccine effects, such as the mean reduction in susceptibility and the mean reduction in the potential to infect others, per infectious contact, can be estimated with precision. However, a more specific parameter reflecting the reduction in infectivity for individuals partially responding to vaccination is not estimated well in the application. An evaluation of the method using artificial data shows that this parameter can be estimated with greater precision when we have outbreak data on a large number of small households.

KEY WORDS: Design of vaccine trials; Household outbreaks; Infectivity; Smallpox; Susceptibility to infection; Transmission of infection; Vaccine efficacy.

1. Introduction

For the control of vaccine-preventable diseases, it is important to have measures of the effectiveness of the vaccine and estimates of these measures. Various such measures have been considered in the literature. Smith, Rodrigues, and Fine (1984) give two forms for the way vaccination might affect an individual's susceptibility to infection. Halloran, Haber, and Longini (1992) consider responses to vaccination that include the types of response considered by Smith et al. (1984), some of which allow for the possibility that vaccination reduces the vaccinated individual's infectivity should he become infected. Here, we consider a family of responses to vaccination that is wider in terms of effects on susceptibility and infectivity. Our focus is on estimating parameters of this family of vaccination responses that indicate how effective the vaccine is at reducing disease transmission. There is particular interest in estimating how much vaccination reduces infectivity. This reduction is difficult to estimate, because it requires data on the source of infection; see Rhodes, Halloran, and Longini (1996). While we can observe who gets infected, there is often little information on who is responsible for a particular infection. This prompts us to consider data on household outbreaks, because such data contain some implicit information on who is likely to have infected whom.

Most of the current literature concerned with how much vaccination reduces infectivity is motivated by potential vaccines for reducing HIV infection and focuses on vaccine trials involving singles and couples, which can be viewed as small households; see Koopman and Little (1995), Rida (1996), Datta, Halloran, and Longini (1999) and Longini et al. (1998). The present approach differs from these by considering an application to outbreaks in a range of household sizes and by having either a different or a more general formulation for the vaccine response. Longini et al. (1998) propose an approach to estimating reduced infectivity that is not based on household-outbreak data. Instead, they use epidemic data on multiple populations consisting of uniformly-mixing individuals, where the populations differ in their vaccination coverage.

The layout of the present article is as follows. In Section 2, we describe a model for household outbreaks of an infectious disease transmitted from person to person. Although our application assumes homogeneity of unvaccinated individuals, our analysis requires a model for different kinds of individual, because individuals may differ in their vaccine response, so that vaccination generates a community of heterogeneous individuals. Infection may be transmitted to individuals either by sources outside the household or by an infected household member.

In Section 3, the class of vaccination response formulated by Becker and Starczak (1998) is described, and we introduce some measures of vaccination effect for this class. We then specify the subclass of vaccination responses that is used in the proposed analysis. Our subclass of responses allows for the possibility of vaccine failure, complete immunity, and partial immunity, with the last response including a reduction in both susceptibility and infectivity.

The probability distribution for the eventual number of vaccinated and unvaccinated cases in a given household is derived in Section 4. This distribution can be derived using the results of Section 2, essentially by conditioning on the vaccine responses of individuals. Section 5 briefly describes the form of the data and the proposed method for making inferences. A Bayesian approach is proposed, because it seems best able to deal with inferences about this complex model, bearing in mind that the transmission model is specified by recursive equations and that the flexible description for vaccine response introduces many parameters. In Section 6, the methods are illustrated by an application to data from a smallpox epidemic in a partially immunized community. In Section 7, the methods are evaluated further using artificial data. The article concludes with some discussion.

2. A Transmission Model for Household Outbreaks

Our focus is on a community of individuals who reside in households. Some of the individuals have been vaccinated. It is assumed that prior to vaccination, all individuals are equally susceptible to the disease in question, and equally infectious if infected. However, variability in vaccination response induces differences among the individuals. Therefore, the inferences need a multitype model for transmission of infection. The following model, from Addy, Longini, and Haber (1991), assumes that individuals may be infected either from the community at large, or by an infected household member.

Consider a population consisting of individuals who reside in households of various sizes. Individuals are of k different types, labeled $1, \ldots, k$. Each individual of type i independently avoids infection from the community, during the observation period $(0, \tau)$, with probability q_i. Here, we implicitly assume that $q_i = \mathrm{e}^{-\int_0^\tau \eta_i(x)dx}$, where $\eta_i(x)$ is the intensity of infection exerted at time x on a type-i individual by infectious individuals outside the household. The observation period may be any interval that includes an entire epidemic. However, the model also applies when $(0, \tau)$ contains only part of an epidemic or when transmission is endemic, although then the observation period must be large relative to the duration of a typical infectious period for an infected individual. An infected type-i individual is assumed to be infectious for a period T_i, where T_i is random with moment generating function $\phi_i(s) = \mathrm{E}(\mathrm{e}^{-sT_i})$. During his infectious period, a type-i individual has infectious contacts with a given type-j household member at the points of a Poisson process of rate β_{ij}. Any such contact results in infection of the type-j individual. The Poisson processes describing infections between different infective-susceptible pairs are assumed to be independent. It follows that, conditional upon T_i, a type-i individual infects each given type-j susceptible household member independently with equal probability $1 - \mathrm{e}^{-\beta_{ij}T_i}$. Removing

the conditioning yields that the probability of a given type-j susceptible avoiding infection from a given type-i infective is $\mathrm{E}(\mathrm{e}^{-\beta_{ij}T_i}) = \phi_i(\beta_{ij})$, although now the fates of different individuals are not independent, in general. Note that to avoid infection, an individual must fail to be infected both from the community, and also from all infectives within the household.

The probability distribution for the final outcome of the transmission model described above can be expressed in terms of the above avoidance probabilities. We use bold letters to denote k-dimensional vectors and define

$$\binom{\mathbf{s}}{\mathbf{j}} = \binom{s_1}{j_1} \cdots \binom{s_k}{j_k}, \qquad \mathbf{s}^{\mathbf{j}} = \prod_i s_i^{j_i},$$

$$\text{and} \qquad \sum_{\mathbf{s} < \mathbf{j}} = \sum_{s_1 \le j_1} \cdots \sum_{s_k \le j_k},$$

where at least one inequality in the sums on the right of the last equation must be strict.

Let $p_{\mathbf{s},\mathbf{r}}(\mathbf{j})$ denote the probability that a household containing $\mathbf{s}$ initial susceptibles (i.e., s_1 of type 1, s_2 of type 2, etc.) and $\mathbf{r}$ initial infectives has $\mathbf{j}$ additional infections during the observation period. Following Addy et al. (1991), the final-size probabilities can be derived recursively from the equations

$$p_{\mathbf{s},\mathbf{r}}(\mathbf{j}) = \binom{\mathbf{s}}{\mathbf{j}} \prod_i \phi_i \left\{ \sum_k (s_k - j_k)\beta_{ik} \right\}^{j_i - r_i} \mathbf{q}^{\mathbf{s}-\mathbf{j}}$$

$$- \sum_{\mathbf{w} < \mathbf{j}} \binom{\mathbf{s} - \mathbf{w}}{\mathbf{j} - \mathbf{w}} p_{\mathbf{s},\mathbf{r}}(\mathbf{w}) \prod_i \phi_i \left\{ \sum_k (s_k - j_k)\beta_{ik} \right\}^{j_i - w_i},$$

$$\mathbf{0} \le \mathbf{j} \le \mathbf{s}. \quad (1)$$

In the absence of vaccination, we assume that the population is described by a single-type model, obtained by setting $k = 1$, $q_1 = q$, $T_1 = T$, and $\beta_{11} = 1$. The last assumption can be made without loss of generality, since β_{11} and T cannot be separately identified from final outcome data.

3. Vaccination

The purpose of vaccinating an individual is to protect the individual against infection and, more generally, to reduce transmission within the community. Our interest therefore lies in the effect of vaccination on an individual's susceptibility to infection and his subsequent potential to infect others. Consequently, it is natural to formulate a description of the effects of vaccination in terms of changes in the forces of infection exerted on individuals and by individuals. The framework we now describe was first proposed by Becker and Starczak (1998), who used it to study the effect of variable vaccination response on the fraction who need to be vaccinated to prevent major epidemics. Here the focus is on measures of vaccine efficacy and their estimation.

To formulate the effect of vaccination on susceptibility, consider a susceptible individual who is subjected to the intensity of infection λ_t, at time $t \ge 0$. Here λ_t is the sum of the intensities exerted by all infectious household members and from the community at large. The probability that the individual is infected in the time increment $[t, t + dt)$ is $\lambda_t dt + o(dt)$ if he has not been vaccinated and $A\lambda_t dt + o(dt)$ if he has been

vaccinated. The nonnegative quantity A is the proportionate reduction in the probability that a vaccinated individual is infected, per contact, relative to an unvaccinated individual. We assume that A is random, thus allowing susceptibility to vary between vaccinated individuals. This is appropriate, because sometimes vaccines fail and there are differences in the immune systems of individuals.

Individuals who have an immunological response to vaccination may have a milder form of the illness, if ever infected, making them less likely to transmit the infection to others. To quantify a possible reduction in the potential to infect others, let B denote the proportionate reduction in the mean number of infectious contacts that a vaccinated individual has, compared to the number he would have if he were not vaccinated, given that he has been infected. The relative infection potential B lies between zero and unity when vaccination reduces infectivity. By assuming that B is random, we can accommodate variation in those immunological responses that affect infectivity.

The effects of vaccination on susceptibility and infectivity are described by the joint probability distribution of A and B,

$$\Pr(A = a_i, B = b_i) = p_i, \qquad i = 1, 2, \ldots, k. \qquad (2)$$

This general formulation can provide a good approximation to any plausible vaccine response. It is of interest to estimate certain characteristics of this distribution, particularly ones that indicate how effective the vaccine is. Halloran, Longini, and Struchiner (1999) review a number of ways to measure the efficacy of a vaccine. Here, we consider these for a larger family of vaccine response. The quantity $VE_S = 1 - \mathrm{E}(A)$ measures the reduction in susceptibility induced by the vaccine. Also relevant are $\Pr(A = 0)$, the probability of acquiring complete protection against infection and $\Pr(A = 1)$, the probability of acquiring no protection. Measures of infectivity reduction need to be constructed with some care. The quantity $VE_I = 1 - \mathrm{E}(B)$ seems to be an appropriate measure of the reduction in infectivity, but it is not well defined when $\Pr(A = 0) > 0$, as is hopefully the case for some vaccines. Whenever the value of A for a vaccinated individual is zero, that individual cannot be infected, so his value of B has no meaning. To overcome this concern, one might judge the vaccine effect on infectivity by considering both the probability $\Pr(A = 0) = c$ and the quantity $1 - \mathrm{E}(B \mid A > 0)$. Alternatively, one might use the quantity $VE_{IS} = 1 - \mathrm{E}(AB)$, which measures the reduction due to vaccination to the extent to which a susceptible individual can both become infected and then infect others. This measure is not affected by the definition of B when $A = 0$. Furthermore, VE_{IS} also plays a central role in the extent of transmission within a partially vaccinated community; see Becker and Starczak (1998).

To incorporate the vaccine effect (A, B) into the transmission model, we begin with homogeneous, unvaccinated individuals. Following vaccination of some individuals, the infection probabilities involving contacts with vaccinated individuals are specified by the following. A vaccinated individual with relative susceptibility $A = a$ has probability $q^a = \mathrm{e}^{-a \int_0^t \eta(x)dx}$ of avoiding infection from the community, where $\eta(x)$ is the intensity of infection exerted at time x from individuals outside

Table 1
Parameters used to describe vaccine effects in our application

f	probability of vaccine failure
c	probability that vaccine gives complete protection
a	relative susceptibility if the vaccine is partially effective
b	relative infectivity potential if the vaccine is partially effective

the household. For a randomly selected individual, or unconditionally, the probability of avoiding infection from the community is $\mathrm{E}(q^A)$. The unconditional probability of a vaccinee avoiding infection from an infected unvaccinated household member is $\mathrm{E}(\mathrm{e}^{-AT}) = \mathrm{E}\{\phi(A)\}$, using the fact that A and T are associated with different individuals and therefore independent. Conversely, the probability that an unvaccinated individual avoids infection from a vaccinated, infected household member is $\mathrm{E}(\mathrm{e}^{-BT}) = \mathrm{E}\{\phi(B)\}$, where it is assumed that B and T are independent. Finally, the probability that a susceptible vaccinee avoids infection when exposed to an infected vaccinee of the same household is $\mathrm{E}(\mathrm{e}^{-ABT}) = \mathrm{E}\{\phi(AB)\}$, where A and B are associated with different individuals and therefore independent.

In our application, we restrict attention to a form of the vaccine response distribution (2) with three distinct outcomes. The vaccination fails completely with probability $f = \Pr(A = 1, B = 1)$, it induces complete immunity with probability $c = \Pr(A = 0)$, or it induces partial immunity, which we specify by $\Pr(A = a, B = b) = 1 - f - c$, where $0 < a, b < 1$. This model for the vaccine response has four parameters, summarized in Table 1. It is of interest to see how well they, and the measures VE_S and VE_{IS}, can be estimated from household outbreak data.

4. Final-Size Probabilities for a Partly Vaccinated Household

The outcome data used to make inferences about model parameters consist, for each household, of the eventual number of cases among vaccinated individuals and the eventual number of cases among unvaccinated individuals. To construct the likelihood function corresponding to such data, we need an expression for the probability

$$\pi_{u,v}(i, j) = \Pr(i \text{ out of } u \text{ unvaccinated and } j \text{ out of}$$

$$v \text{ vaccinated become infected})$$

for a household initially comprised of u unvaccinated and v vaccinated susceptibles, where $0 \leq i \leq u$ and $0 \leq j \leq v$. While these probabilities are not given directly by the final-size probabilities of Section 2, they can be evaluated from these probabilities. For these calculations, it is sufficient to consider a model with only two types of individual, even though there are unvaccinated individuals and three possible vaccine responses. This is because individuals with a failed vaccination are the same "type" as unvaccinated individuals, while vaccinees who acquire complete immunity play no part in the transmission of infection.

Consider, therefore, a two-type form of the model in Section 2, where type-1 individuals are those who are either unvaccinated or vaccinated with a failure response, and

Table 2
Probability of avoiding infection

	Type-1 infective	Type-2 infective
Type-1 susceptible	$\mathrm{E}(\mathrm{e}^{-T}) = \phi(1)$	$\mathrm{E}(\mathrm{e}^{-bT}) = \phi(b)$
Type-2 susceptible	$\mathrm{E}(\mathrm{e}^{-aT}) = \phi(a)$	$\mathrm{E}(\mathrm{e}^{-abT}) = \phi(ab)$

type-2 individuals are vaccinees who acquired partial immunity. Type-1 individuals are infected from outside the household, during the observation period $(0, \tau)$, with probability $q_1 = q$, and type-2 individuals with probability $q_2 = q^a$. The probability that a type-1 individual avoids infection from another infected type-1 household member is $\mathrm{E}(\mathrm{e}^{-T}) = \phi(1)$. The four avoidance probabilities corresponding to possible infective-susceptible pairs are summarized in Table 2. In the notation of Section 2, we have $\beta_{11} = 1$, $\beta_{12} = a$, $\beta_{21} = b$ and $\beta_{22} = ba$.

Let $p_{s_1, s_2}(j_1, j_2)$ denote the probability that j_1 out of s_1 initially susceptible type-1 individuals and j_2 of s_2 initially susceptible type-2 individuals become infected. The system of equations (1), for the final-size probabilities, then becomes

$$
p_{s_1, s_2}(j_1, j_2) =
$$

$$
\begin{cases}
q^{s_1 + a s_2}, \quad (j_1, j_2) = (0, 0), \\
\binom{s_1}{j_1}\binom{s_2}{j_2} q^{s_1 - j_1 + a(s_2 - j_2)} \phi\{s_1 - j_1 + (s_2 - j_2)a\}^{j_1} \\
\quad \times \phi\{(s_1 - j_1)b + (s_2 - j_2)ba\}^{j_2} \\
- \sum_{r_1=0}^{j_1} \sum_{r_2=0}^{j_2-1} \binom{s_1 - r_1}{j_1 - r_1}\binom{s_2 - r_2}{j_2 - r_2} p_{s_1, s_2}(r_1, r_2) \\
\quad \times \phi\{s_1 - j_1 + (s_2 - j_2)a\}^{j_1 - r_1} \\
\quad \times \phi\{(s_1 - j_1)b + (s_2 - j_2)ba\}^{j_2 - r_2} \\
- \sum_{r_1=0}^{j_1-1} \binom{s_1 - r_1}{j_1 - r_1} p_{s_1, s_2}(r_1, j_2) \phi\{s_1 - j_1 + (s_2 - j_2)a\}^{j_1 - r_1}, \\
\hspace{6cm} (j_1, j_2) \neq (0, 0).
\end{cases}
$$

The probabilities, for given s_1 and s_2, can be obtained by sequentially computing the probabilities for $(1, 0), \ldots, (s_1, 0)$, $(0, 1), \ldots, (s_1, 1), \ldots, (s_1, s_2)$ using the above equations.

The final outcome probabilities $\pi_{u,v}(i, j)$ for households with u unvaccinated and v vaccinated members, none of whom were previously infected, can be obtained from the $\{p_{s_1, s_2}(j_1, j_2)\}$ by

$$
\pi_{u,v}(i, j) = \sum_{r,s,t; r+s+t=v} \begin{pmatrix} v \\ r \ \ s \ \ t \end{pmatrix} f^r c^s (1 - f - c)^t
$$

$$
\times \sum_{k=0}^{\min(j,r)} \frac{\binom{u}{i}\binom{r}{k}}{\binom{u+r}{i+k}} p_{u+r,t}(i+k, j-k).
$$

The factor after the first summation sign is the trinomial coefficient $v!/(r!s!t!)$, where $r + s + t = v$. This formula is obtained by conditioning on the vaccine responses among the v vaccinated individuals and summing over those configurations with j vaccinated members infected (regardless of their responses). The hypergeometric factor in front of $p_{u+r,t}(i+k, j-k)$ appears because there are $u + r$ fully susceptible individuals, including u unvaccinated individuals and r members whose vaccination failed; we require the probability that exactly i of the $i + k$ infected are unvaccinated.

To compute $\pi_{u,v}(i, j)$ for $0 \leq i \leq u$, $0 \leq j \leq v$, it is necessary to find $p_{s_1, s_2}(j_1, j_2)$ for $s_1 = u, \ldots, u+v$ and $s_2 = 0, \ldots, v - (n_1 - u)$. In practice, it is expeditious to compute the relevant $p_{s_1, s_2}(j_1, j_2)$ terms in advance, since they are required for several different $\pi_{u,v}(i, j)$ terms.

5. Form of the Data and Proposed Inferences

Assume that households are sampled from the community at random, in a way that may depend on the size of the household and/or on the number of vaccinated and unvaccinated household members, but may not depend on the number of cases there are, or will be, in the household. Data on the number of members infected in these households, together with the vaccination status of household members and cases, are to be used to make inferences. In the count of household members, individuals who are known to have acquired immunity from a previous infection are excluded. The same observation period $(0, \tau)$ applies to all households. During the observation period there must be a sufficiently large cumulative force of infection; otherwise, the number of cases will be too small to make useful inferences. Nevertheless, the sample of households will generally include some that have no cases.

The data set has the form

$$
\mathbf{n} = \{n_{u,v}(i, j) : 0 \leq i \leq u \leq u_{\max}, 0 \leq j \leq v \leq v_{\max}\},
$$

where $n_{u,v}(i, j)$ is the number of households in which i out of u unvaccinated and j out of v vaccinated individuals were infected, and $u_{\max}$ and $v_{\max}$, respectively, denote the largest numbers of unvaccinated and vaccinated individuals present in households.

The likelihood corresponding to this data set is given by

$$
L(a, b, c, f, q, \mu) = \prod_{0 \leq i \leq u \leq u_{\max}} \prod_{0 \leq j \leq v \leq v_{\max}} \{\pi_{u,v}(i, j)\}^{n_{u,v}(i,j)},
\tag{3}
$$

where μ denotes the (possibly one-dimensional) parameter vector of the distribution of the infectious period T. The parameters c, f, q are probabilities and so have domain $[0, 1]$. The parameters a and b are positive and no greater than 1 if vaccination does not increase susceptibility or infectivity, while the domain of μ depends on the parameterization used for the distribution of T.

We now consider Bayesian inference for the model parameters a, b, c, f, q, and μ given the data $\mathbf{n}$, by investigating the posterior density $\varphi(a, b, c, f, q, \mu \mid \mathbf{n})$. By Bayes' Theorem, we have

$$
\varphi(a, b, c, f, q, \mu \mid \mathbf{n}) \propto L(a, b, c, f, q, \mu)\varphi(a, b, c, f, q, \mu),
\tag{4}
$$

where L is defined by (3) and $\varphi(a, b, c, f, q, \mu)$ is the prior density on the parameter vector (a, b, c, f, q, μ).

To investigate the posterior density given by (4), we use a Markov chain Monte Carlo (MCMC) algorithm; see Gilks, Richardson, and Spiegelhalter (1996) for an introduction to MCMC methods, and O'Neill et al. (2000) for applications to epidemic models. Specifically, we use a single-component Hastings algorithm. The updating of a single parameter χ (so that χ is one of a, b, etc.) with current value θ, say, is performed as follows. A new value, $\tilde{\theta}$, is drawn from a proposal

density $g_\chi(\tilde{\theta}\,|\,\theta)$. The new value is accepted with probability

$$\min\left\{\frac{\varphi(\ldots,\tilde{\theta},\ldots)}{\varphi(\ldots,\theta,\ldots)}\frac{g_\chi(\theta\,|\,\tilde{\theta})}{g_\chi(\tilde{\theta}\,|\,\theta)},1\right\}.$$

If $\tilde{\theta}$ is not accepted, then χ remains as θ. Note that, although the choice of g_χ is essentially arbitrary, the particular distribution used affects the convergence properties of the algorithm.

6. Application

To illustrate the proposed method for making inferences, we use data on observed household outbreaks of smallpox. The data set is well suited for this illustration, because it is essentially of the form needed and is of a size that can be managed in a vaccine trial. The properties of the vaccine are also suitable. Specifically, the vaccine is not transmissible and it is designed for an antigenically stable pathogen.

6.1 *The Data*

The data are taken from an epidemic of *variola minor,* the mild form of smallpox, that occurred in Bragança Paulista County (Brazil) in 1956; see Angulo (1976). The households were not selected in strict accordance with the sampling scheme assumed in Section 5 but, for our analysis, we assume that the households were sampled without reference to the number of cases. There are a total of 338 households, including 126 without any cases. The household size ranges from 1 to 12 individuals, with a mean household size of 4.56. In total, there are 1542 individuals, of whom 809 were vaccinated and 733 unvaccinated. During the epidemic, 425 of the unvaccinated individuals and 85 of the vaccinated individuals were infected. The data on households affected by smallpox were previously used by Becker and Angulo (1981) to compare the susceptibility of vaccinated and unvaccinated individuals. They estimated parameters of a two-type chain binomial model using maximum likelihood methods, and also fitted a continuous-time model by using martingale estimating equations.

6.2 *Implementation Details*

We present two different analyses corresponding to two models, labeled 1 and 2. In model 1, the infectious period T is nonrandom, while in model 2, T is assumed to have an exponential distribution. The two models thus represent different assumptions regarding the within-household infection process. Comparison of the inferences drawn using these two models enables us to assess informally the degree to which estimation for vaccine efficacy is robust with respect to the assumptions regarding disease transmission. The details are as follows.

For model 1, T is assumed to be a constant with moment generating function $\phi(s)=\mu^s$, so that in the absence of any vaccination, μ is the probability that a given susceptible avoids infection from a given infective household member. The assumption that T is constant is known as the Reed-Frost assumption; see Andersson and Britton (2000). Each of the six model parameters was assigned an independent Uniform(0,1) prior density. Within the Hastings algorithm, each of the proposal densities g was set as Uniform(0,1).

For model 2, T is assumed to have an exponential distribution with mean λ^{-1}. This is known as the general stochastic epidemic assumption; see Andersson and Britton (2000). For convenience, within the MCMC algorithm, the reparameterisation $\nu=\mathrm{e}^{-\lambda}$ was used, so that $\nu\in(0,1)$. Prior distributions and proposal distributions were the same as for model 1.

6.3 *Results*

Posterior density summaries are given for model 1 in Table 3 and for model 2 in Table 4. The quantity $\mu=\lambda\,/(\lambda+1)$ in Table 4 is the avoidance probability between an unvaccinated infective-susceptible household pair, and so has the same interpretation as the μ of Table 3. In both tables, the values are based on samples of size 10^4 from the MCMC output.

The characteristics of the posterior distributions given in the two tables are very similar, which indicates that the results are not sensitive to substantially different assumptions about variation in the infectious period. Consequently, the remarks below apply equally to either model. Additionally, the plots in Figures 1 and 2, both taken from model 1, are very similar to the corresponding plots for model 2 (not shown).

The vaccine failure probability, $f=\Pr(A=1,B=1)$, is a quantity of major interest. For these data, f is estimated with precision that is of practical value. Specifically, the data are not compatible with a failure rate that is greater than 20%.

The probability of acquiring complete immunity, $c=\Pr(A=0)$, is also a quantity of major interest. Its estimation is not very precise, although the results indicate that the data are not compatible with values of c greater than 80%.

Inference for a also lacks precision. Indeed, inferences for a, the relative susceptibility of partially immune vaccinees,

Table 3
Posterior density summary, model 1

	a	b	c	f
Mean	0.217	0.589	0.440	0.078
Median	0.163	0.623	0.460	0.072
S. dev	0.166	0.264	0.228	0.051
95% C.I.	(0.054,0.572)	(0.106,0.959)	(0.055,0.776)	(0.007,0.170)
	q	μ	E(A)	E(AB)
Mean	0.629	0.678	0.157	0.127
Median	0.629	0.678	0.155	0.127
S. dev	0.022	0.027	0.025	0.035
95% C.I.	(0.591,0.665)	(0.633,0.721)	(0.118,0.201)	(0.071,0.186)

<div align="center">

Table 4
Posterior density summary, model 2

</div>

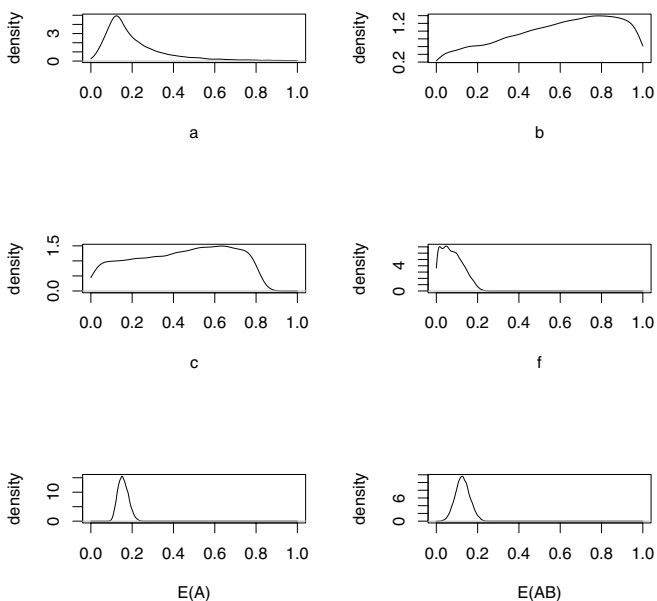|	a	b	c	f
Mean	0.198	0.560	0.438	0.070
Median	0.148	0.576	0.452	0.063
S. dev	0.156	0.264	0.228	0.048
95% C.I.	(0.053,0.526)	(0.100,0.954)	(0.054,0.775)	(0.006,0.160)
	q	μ	E(A)	E(AB)
Mean	0.619	0.651	0.143	0.113
Median	0.619	0.652	0.141	0.112
S. dev	0.022	0.031	0.026	0.037
95% C.I.	(0.582,0.655)	(0.599,0.701)	(0.105,0.189)	(0.054,0.176)

and c, the probability of complete immunity, are closely tied together, as is revealed by the scatter plot for the values of c and a in Figure 2. First, note that the lack of points in the bottom right-hand corner of Figure 2 indicates that the vaccine is effective at reducing susceptibility, because the data are not compatible with c being small and a being large simultaneously. With c large, the data are compatible with most values of a. Conversely, with a small, the data are compatible with most values of c. Specifically, the data are not able to distinguish between $(a, c) \approx (0, 0)$ and $(a, c) \approx (1, 0.8)$, which both indicate low susceptibility.

For a partial-immunity response, the reduction in infectivity is indicated by the value of b. In this application, the credible interval for b is rather wide. This result is disappointing in view of our hope that household data provide a way to make inferences about reduced infectivity. Note, however, that the marginal posterior density of b, shown in Figure 1,

has a slight peak. More specifically, it is not simply uniform. Thus, the data do provide some, albeit not much, information about b.

The inferences regarding the baseline parameters, q and μ, of the transmission model are reasonably precise. Information about q comes primarily from the fact that 502 individuals are known to escape transmission from outside their household; this large number explains the precision in q. The precision of inference about μ arises because the number of within-household exposures between pairs is large.

The parameters c, f, q, and μ are probabilities, making a Uniform(0,1) prior distribution a natural choice. On the other hand, using a Uniform(0,1) prior distribution for a and b, as above, assumes that the vaccine reduces susceptibility and infectivity. This seems reasonable for the smallpox vaccine, but vaccines with adverse effects have been reported; see Kim et al. (1969). Adverse effects are of particular concern for sexually transmitted diseases, since being vaccinated might influence risk-taking behavior. We applied the method of inference with a Uniform(0,3) prior distribution for a and b. Inferences for c, f, q, μ and E(A) were minimally affected. The upper credible bound for a slipped out to 0.94, while inference for b were affected substantially (95% C.I. (0.21,2.7)). In short, the information about b in these data is not sufficient to dominate its prior distribution.

7. Evaluating the Method for Making Inferences

It is desirable to evaluate the performance of such inferences by conducting a comprehensive study, based on many simulated data sets for each of many points in the six-dimensional parameter space. This is impracticable, because the method involves very intensive computation for each chosen set of parameter values and each data set. Here, we conduct a very limited evaluation, from which we hope to learn something about (a) how well reduced infectivity can be estimated from data on outbreaks in smaller households, (b) how much inferences are affected by the way vaccinated individuals are distributed over the households, and (c) how precision of inferences improves as the sample size increases.

We choose just one set of parameter values, namely, $a = 0.4$, $b = 0.5$, $c = 0.45$, and $f = 0.05$, giving E(A) = 0.25 and E(AB) = 0.15, together with $q = 0.6$ and $\mu = 0.7$. These values are chosen because they reflect a plausible vaccine response
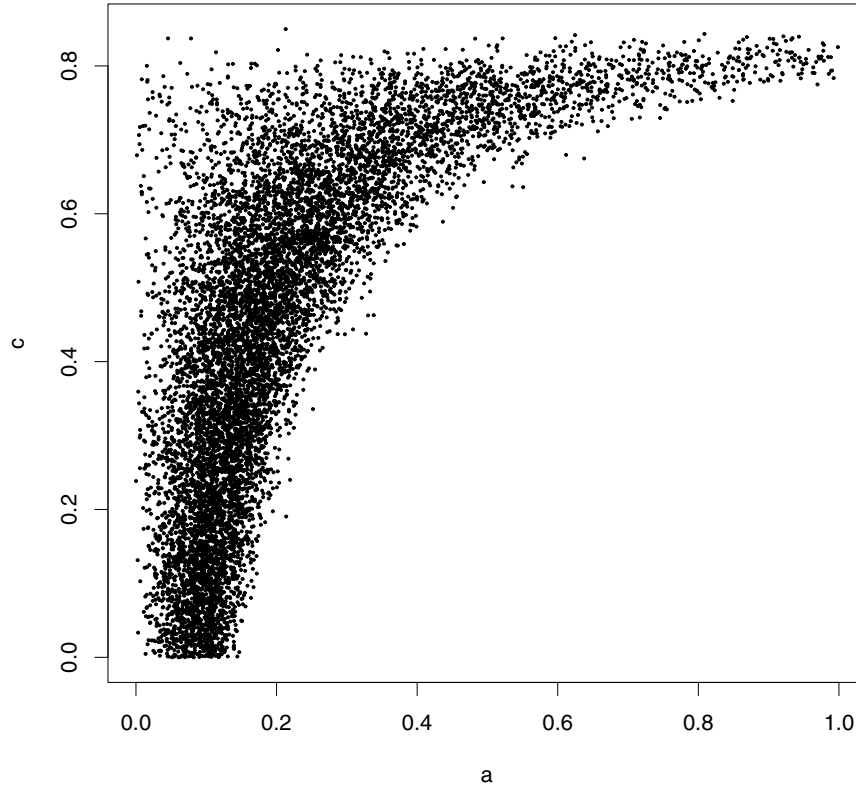

Figure 1. Posterior density plots for a (relative susceptibility), b (relative infectivity potential), c (complete protection probability), f (vaccine failure probability), E(A), and E(AB), model 1.

Figure 2. Scatterplot of *a* and *c*, model 1.

and a plausible amount of transmitted infection within households.

We consider four study designs. In the first two designs, all households are of size two, and they are further specified by the total number of households and the distribution of vaccination over the households. Specifically, we have:

Design 2a All households have exactly one vaccinated and one unvaccinated member, and

Design 2b Half the households have both members vaccinated and the other half have both members unvaccinated.

The remaining two designs have all households of size four and are of the following two types:

Design 4a All households have exactly two vaccinated and two unvaccinated members, and

Design 4b Half the households have four vaccinated members and the other half have four unvaccinated members.

Instead of simulating many data sets for each design, we make inferences from just one data set for each design in which the observed frequencies of different household outbreaks are equal to the expected frequencies. In other words, we simply take the integer part of the expected frequencies for a given total number of households. In a sense, we then have the ideal data set for those parameter values; we are interested in seeing if the method of inference can identify the parameters when the number of households is of a size that is manageable in practice. We computed posterior distributions for data sets based on 250, 500, and 1000 households. An illustration of the results is given in Table 5.

Estimation of *a* was usually reasonable in all designs considered. We found that smaller household sizes do indeed provide greater precision for the estimation of *b*, in the sense that the mean posterior values are closer to the true value $b = 0.5$. This is consistent with results found by Rida (1996) and Datta et al. (1999) in related settings. Additionally, for *b*, designs 2a and 4a proved better than 2b and 4b, respectively; one explanation is that in households where all

Table 5
Posterior density summary for the evaluation data set (using design 4b with a total of 500 households)

	a	b	c	f	q	μ	E(A)	E(AB)
True value	0.40	0.50	0.45	0.05	0.60	0.70	0.25	0.15
Posterior mean	0.41	0.45	0.54	0.14	0.60	0.70	0.24	0.19
Posterior s. dev	0.27	0.28	0.20	0.08	0.03	0.03	0.04	0.06

individuals are vaccinated, it is relatively unlikely that there are many infections at all, and thus there is limited information for inference about b (e.g., in design 2b, only 18 out of 250 households contained vaccinees who became infected).

Estimation of $\mathrm{E}(A)$ was generally good across all designs, while that for $\mathrm{E}(AB)$ varied considerably. The most accurate posterior estimate came from design 4b. In fact, since $\mathrm{E}(AB) = f + ab(1 - f - c)$ for typical values of a, b, c it is clear that f dominates the estimation of $\mathrm{E}(AB)$. However, estimating f is made complicated by the fact that the actual value of $f = 0.05$ is quite close to zero, and so the data do not contain much information about f. This was reflected in the results; as sample size increased, so the estimation for f improved. Finally, in general, there was some increase in precision, in terms of smaller credible intervals, for larger sample sizes. However, this increase was typically only of the order of around 5–10%.

8. Discussion

Our aim has been to develop and test a method for making inference about measures of vaccine efficacy given household outbreak data. The modeling framework for vaccine response allows consideration of measures of reduction in susceptibility and potential to transmit infection. We restricted attention to a three-type vaccine response, but could have modeled partial vaccine response more elaborately than a single point mass at (a, b). It seems unlikely that the additional parameters would be identifiable from the data, but summary measures such as $\mathrm{E}(A)$ should still be estimable with similar precision. It is not wholly surprising that estimation of $\mathrm{E}(A)$ is fairly precise (see Figure 1), because it is possible to obtain estimable bounds on $\mathrm{E}(A)$ without specifying the distribution of (A, B), and these bounds can be reasonably tight; see Becker and Utev (2002).

The precise estimation of $VE_{IS} = 1 - \mathrm{E}(AB)$ is notable, particularly in view of the pessimism expressed in Section 2.4 of Becker and Starczak (1998). On the other hand, distinct characteristics of the marginal distribution of B, such as the value b, are only estimated with a precision of practical value when we have data on a large number of small households. Further study of this issue would facilitate the design of vaccine trials that take into account such issues as the number of individuals who should be vaccinated and how these individuals should be distributed over the households; see Rida (1996).

Finally, we have assumed, as is common practice, that the rate of making infectious contacts is constant over the duration of the infectious period. The methods do not require this assumption. If the rate at which individuals shed infectious agent is not constant over the infectious period, then one can proceed similarly by replacing βT, as in Section 2, by the area under the infectiousness function.

Résumé

Cet article propose une méthode pour faire des inférences sur des mesures diverses d'efficacité vaccinale. Ces mesures décrivent des diminutions dans la susceptibilité et dans le potentiel de transmettre l'infection. La méthode utilise des données d'épidémies intra-familiales et se base sur un modèle qui permet la transmission de l'infection à la fois de l'intérieur d'un foyer et de l'extérieur. L'utilisation de données intra-familiales est motivée par l'espoir qu'elles soient informatives sur la diminution du potentiel de transmission de l'infection, induite par le vaccin, car les épidémies intra-familiales contiennent des informations sur la source possible d'infection. A des fins d'illustration, la méthode est appliquée à des données d'épidémies intra-familiales de variole. Ces données sont dans la forme souhaitée et le nombre de foyers est d'une taille qui permet de gérer un essai vaccinal. Il a été montré que les effets de la vaccination tels que la diminution moyenne de la susceptibilité et la diminution moyenne du potentiel d'infecter les autres, par contact infectieux, peut être estimé avec précision. Cependant, un paramètre plus spécifique reflétant la réduction de l'infectiosité pour des individus répondant partiellement à la vaccination n'est pas bien estimé dans l'application. Une évaluation de la méthode en utilisant des données artificielles montre que ce paramètre peut être estimé avec une plus grande précision lorsque l'on dispose de données épidémiques sur un grand nombre de petits foyers.

References

Addy, C. L., Longini, I. M., and Haber, M. (1991). A generalized stochastic model for the analysis of infectious disease final size data. *Biometrics* **47,** 961–974.

Andersson, H. and Britton, T. (2000). *Stochastic Epidemic Models and Their Statistical Analysis.* Lecture Notes in Statistics **151**. New York: Springer.

Angulo, J. J. (1976). Variola minor in Bragança Paulista County, 1956: Overall description of the epidemic and of its study. *International Journal of Epidemiology* **5,** 359–366.

Becker, N. G. and Angulo, J. J. (1981). On estimating the contagiousness of a disease transmitted from person to person. *Mathematical Biosciences* **54,** 137–154.

Becker, N. G. and Starczak, D. N. (1998). The effect of random vaccine response to prevent epidemics. *Mathematical Biosciences* **154,** 117–135.

Becker, N. G. and Utev, S. (2002). Protective vaccine efficacy when vaccine response is random. *Biometrical Journal* **44,** 29–42.

Datta, S., Halloran, M. E., and Longini, Jr., I. M. (1999). Efficiency of estimating vaccine efficacy for susceptibility and infectiousness: Randomization by individual versus household. *Biometrics* **55,** 792–798.

Gilks, W. R., Richardson, S., and Spiegelhalter, D. J. (Eds.) (1996). *Markov Chain Monte Carlo in Practice.* London: Chapman and Hall.

Halloran, M. E., Haber, M., and Longini, I. M. (1992). Interpretation and estimation of vaccine efficacy under heterogeneity. *American Journal of Epidemiology* **136,** 328–343.

Halloran, M. E., Longini, I. M., and Struchiner, C. J. (1999). Design and interpretation of vaccine field studies. *Epidemiologic Reviews* **21,** 73–88.

Kim, H. W., Canchola, J. G., Brandt, C. D., Pyles, G., Chanock, R. M., Jensen, K., and Parrott, R. H. (1969). Respiratory syncytial virus disease in infants despite prior administration of antigenic in activated vaccine. *American Journal of Epidemiology* **89,** 422–434.

Koopman, J. S. and Little, R. J. (1995). Assessing HIV vaccine effects. *American Journal of Epidemiology* **142,** 1113–1120.

Longini, I. M., Sagatelian, K., Rida, W. N., and Halloran, M. E. (1998). Optimal vaccine trial design when estimating vaccine efficacy for susceptibility and infectiousness from multiple populations. *Statistics in Medicine* **17,** 1121–1136.

O'Neill, P. D., Balding, D. J., Becker, N. G., Eerola, M., and Mollison, D. (2000). Analyses of infectious disease data from household outbreaks by Markov chain Monte Carlo methods. *Applied Statistics* **49,** 517–542.

Rhodes, P. H., Halloran, M. E., and Longini, I. M. (1996). Counting process models for infectious disease data: Distinguishing exposure to infection from susceptibility. *Journal of the Royal Statistical Society, Series B* **58,** 751–762.

Rida, W. N. (1996). Assessing the effect of HIV vaccination on infectiousness. *Statistics in Medicine* **15,** 2393–2404.

Smith, P. G., Rodrigues, L. C., and Fine, P. E. M. (1984). Assessment of the protective efficacy of vaccines against common diseases using case-control and cohort studies. *International Journal of Epidemiology* **13,** 87–96.