Estimating Vaccine Efficacy From Secondary Attack Rates

M. Elizabeth HALLORAN, Marie-Pierre PRÉZIOSI, and Haitao CHU

Epidemiologists have used secondary attack rates (SARs) to estimate the protective effects of vaccination since the 1930s. SARs can also be used to estimate the effect of vaccination on reducing infectiousness in breakthrough cases. The conventional SAR approach has been to pool the denominators and numerators across transmission units, then to use a confidence interval for a simple relative risk. We demonstrate appropriate model-based methods to estimate vaccine efficacy (VE) from SARs using generalized estimating equations taking correlation within transmission units into account. The model-based procedures require transformation of the parameter estimates to the SAR scale to obtain vaccine efficacy estimates. Appropriate confidence intervals are then based on the bootstrap, with resampling done by transmission unit. We show that the usual confidence intervals are too narrow. We estimated the effect of pertussis vaccination on person-to-person transmission. The results show that pertussis vaccination reduces the ability of a breakthrough clinical case to produce other clinical cases. The methods can be used in evaluating VE for susceptibility and infectiousness from SARs in other infectious diseases.

KEY WORDS: Africa; Bootstrap; Clustering; Efficacy; Generalized estimating equations; Hierarchical models; Pertussis; Random effects models; Secondary attack rate; Vaccine.

1. INTRODUCTION

Vaccine efficacy (VE) measures are usually estimated as $VE = 1 - RR$, where RR is some measure of relative risk in the vaccinated groups compared with unvaccinated groups (Orenstein et al. 1985; Halloran, Struchiner, and Longini 1997). Historically, the protective efficacy of vaccination, VE_S, has been the main interest. The secondary attack rate (SAR), the proportion of susceptibles exposed to an infectious person who become infected, has been used to estimate protective effects of vaccination since the 1930's (Kendrick and Eldering 1939). Recently, the effect of vaccination on reducing infectiousness in breakthrough cases, VE_I, based on the relative SAR from vaccinated cases compared with that from unvaccinated cases, has received more attention (Longini, Datta, and Halloran 1996; Golm, Halloran, and Longini 1998; Halloran et al. 1997). Similarly, the total effect, VE_T, can be estimated from the SAR when both the infective and susceptible are vaccinated compared with the SAR when neither are vaccinated.

Quite often one infectious person exposes several people, possibly within a transmission unit, such as a household. Correlation within transmission units or unmeasured heterogeneity across transmission units could result from, for example, differences in infectivity, differences in mixing within the unit, or genetic variation. The conventional method of estimating VE from SARs fails to take the structure of the clustered binary data into account. Currently, "the total population in each household minus the excluded primary and coprimary cases is added together to get large vaccinated and unvaccinated cohorts. Cases to be included are classified by vaccination status, the attack rates are determined, and vaccine efficacy is calculated" (Orenstein, Bernier, and Hinman 1988). Confidence intervals are based on various methods, none of which takes into account correlation within transmission unit; examples include the methods of Orenstein et al. (1985); Katz, Baptista, Azen, and Pike (1978), used by Fine and Clarkson (1987); Thomas and Gart (1977), used by Storsaeter, Blackwelder, and Hallander (1992); and Francis et al. (1955), used by Kim-Farley et al. (1985).

In this article we present appropriate methods for estimating VE measures based on the SAR that take into account correlation within transmission units, and demonstrate estimation of VE for infectiousness. Controversy still surrounds whether pertussis vaccination reduces circulation of *Bordetella pertussis* (*Bp*), the bacteria that causes pertussis, also known as whooping cough. (Fine and Clarkson 1982, 1987; Miller and Gay 1997; Trollfors et al. 1998; Rohani, Earn, and Grenfell 2000; Taranger et al. 2001; Préziosi et al. 2002). Interested in estimating the effect of pertussis vaccination on transmission, we analyzed data from a population-based study of pertussis in Niakhar, Senegal, using these methods (Préziosi and Halloran 2003).

2. STUDY BACKGROUND

The pertussis study was part of a larger demographic surveillance project conducted in the rural Niakhar region of Senegal (Garenne and Cantrelle 1998; Simondon et al. 1997; Préziosi et al. 2002). The community is composed of Sereer peasant families residing in the 30 villages within the Niakhar study area. Extended families, residing in compounds, were under active longitudinal observation beginning in March 1983, based on annual visits, and from 1987 to 1996, based on weekly visits to each compound. Field workers administered

M. Elizabeth Halloran is Professor, Marie-Pierre Préziosi is Visiting Instructor, and Haitao Chu is a doctoral candidate, Department of Biostatistics, Rollins School of Public Health, Emory University, Atlanta, CA 30322 (E-mail: *mehallo@sph.emory.edu*). Préziosi is also from the Niakhar Project, Institut de Recherche pour le Développement, Dakar, Senegal. This research was partially supported by National Institutes of Health grant R01-AI32042. Préziosi was also partially supported by the French Ministry of Foreign Affairs (grant Lavoisier), the Singer-Polignac Foundation (subvention), and Aventis Pasteur (consultancy fees). Data collection was financed by Institut de Recherche pour le Développement and Aventis Pasteur. The authors are grateful to the Niakhar Project team, in particular Francois Simondon, head of the pertussis clinical trials, and Michel Garenne and Peter Aaby, initiators of pertussis surveillance in the area; and to Coumba Toure Kane and Souleymane Mboup (University Cheikh Anta Diop, Dakar, Senegal), and Nicole Guiso and Isabelle Iteman (Institut Pasteur, Paris, France), in charge of biological surveillance for the pertussis trials. They also thank Bradley Efron for enlightening discussions on bootstrap confidence intervals.

Journal of the American Statistical Association
March 2003, Vol. 98, No. 461, Applications and Case Studies
DOI 10.1198/016214503388619076

structured questionnaires, and experienced physicians checked for illnesses.

Surveillance for pertussis focused on children under age 15 years. A suspected case of pertussis was defined as cough of 8 days duration or longer. All suspected cases and their coresidents living in the same compound were followed actively by a physician. The study was conducted in accordance with the Helsinki Declaration of 1975 (revised 1983) (Préziosi et al. 1997). The usual demographic data, including age, gender, hut, compound, hamlet, and village, were known for each individual in the area. Pertussis vaccination status and dates of vaccination were also known. For each suspected case, the date of symptom onset, duration of cough, type of cough, a wide range of symptoms, results of each biologic diagnostic test done, and physician diagnosis were recorded. This analysis focuses on the calendar year 1993, an epidemic year that produced a large number of cases and extensive exposure to pertussis. The availability of biologic diagnostic tools was excellent. Biologic samples were collected from all consenting suspected cases.

3. METHODS FOR THE SECONDARY ATTACK RATE

3.1 Case Definition

In this article a case of pertussis is defined as requiring clinically, at least 21 days of cough with paroxysms and biologically, either *Bp* isolated from a nasopharyngeal aspirate (bacterio+) or significant increase or decrease in pertussis toxin or filamentous hemagglutinin antibodies as measured by enzyme-linked immunosorbent assay (sero+), or presence of a bacterio+ case within 28 days in the same compound (epilink+) (Préziosi and Halloran 2003). This case definition is similar to that of the World Health Organization (WHO) (1991), but differs slightly in two respects. First, it requires 21 days of cough with some paroxysms, rather than 21 days of continuous paroxysmal cough, the latter being overly specific as stated in a recent WHO meeting (WHO 2001). Second, it allows significant decreases as well as increases in antibodies to accept a positive serology, as detailed elsewhere (Simondon, Iteman, Préziosi, Yam, and Guiso 1998). Indeed, it has been shown in this setting that the serology result could be different between vaccinated and unvaccinated people due to different kinetics of antibody response in these groups.

3.2 Setting Up the Secondary Attack Rate Analysis

We chose the compound as the transmission unit within which it was assumed that susceptibles were exposed to infection by the first case in the unit. The compound is the "home," that is, the residential unit where individuals make privileged contacts and where random mixing is a reasonable assumption. Indeed, for these reasons, the compound is the transmission unit of choice in African rural settings such as here (Garenne, Leroy, Beau, and Sene 1993; Aaby, Samb, Anderson, and Simondon 1996). The first case in a transmission unit is called the index or primary case. A potentially infectious contact, or exposure, was defined as a susceptible living in the same compound during the infectious period of the index case. Exposed susceptibles were children with no history of pertussis living in a compound with an index case. Onset of pertussis symptoms was assumed to be the onset of infectiousness; thus the latent period equals the incubation period. Coprimaries were those cases with onset of cough less than 7 days after that of the index case, assumed to be too soon after the index case to have been infected by the index case. To allow for uncertainty in duration of infectiousness, a secondary case was defined as a case with onset of cough 7 or more days after that of the index case and less than a variable cutoff, specifically no cutoff 56, 42, or 28 days.

Generally, when estimating protective efficacy, VE_S, from SARs, coprimaries are simply ignored in the analysis, entering as neither susceptibles nor infectives (Orenstein et al. 1988; Fine, Clarkson, and Miller 1988). However, the particular interest here was in the effect of vaccine status on infectiousness of the index case. Because primaries and coprimaries often had different vaccine status, compounds with coprimaries were excluded from the analysis.

3.3 Notation

Let n be the number of compounds with a unique index case and m_i be the number of susceptibles in the ith compound. Let y_{ij} be the binary (0, 1) pertussis outcome of the jth susceptible exposed to the index case in the ith compound for any given case definition. Let $\mathbf{x}_{ij} = (x_{ij1}, \ldots, x_{ijp})'$ denote a $p \times 1$ vector of explanatory variables associated with y_{ij}. In particular, let $x_{i\cdot 1}$ denote the vaccine status of the index case in compound i, and x_{ij2} denote the vaccine status of the jth exposed susceptible individual in compound i. Complete pertussis vaccination requires at least three doses of vaccine. In this analysis we consider unvaccinated and fully vaccinated children, with $x_{i\cdot 1} = 0$ for an unvaccinated case and $x_{i\cdot 1} = 1$ for a fully vaccinated index case. Similarly, x_{ij2} is 0 or 1 for the unvaccinated and fully vaccinated susceptibles.

Let N_{vs} be the total number of susceptibles in the n compounds with vaccine status s exposed to index cases with vaccine status v, and a_{vs} be the total number of cases in the N_{vs} susceptibles. In this article, $V, S \in \{0, 3\}$. The subscript "0" denotes unvaccinated, and "3" indicates three doses of vaccine. Additional levels of vaccination are possible, such as $V, S \in \{1, 2\}$ for partially vaccinated people, but are not considered here. The "$\cdot$" subscript represents collapsing over strata. The number of cases and susceptibles in each grouping of interest is

$$a_{vs} = \sum_{i=1}^{n}\sum_{j=1}^{m_i} I_{V=v} I_{S=s} y_{ij}, \qquad N_{vs} = \sum_{i=1}^{n}\sum_{j=1}^{m_i} I_{V=v} I_{S=s},$$

$$a_{\cdot\cdot} = \sum_{i=1}^{n}\sum_{j=1}^{m_i} y_{ij}, \qquad N_{\cdot\cdot} = \sum_{i=1}^{n} m_i,$$

$$a_{\cdot s} = \sum_{i=1}^{n}\sum_{j=1}^{m_i} I_{S=s} y_{ij}, \qquad N_{\cdot s} = \sum_{i=1}^{n}\sum_{j=1}^{m_i} I_{S=s},$$

$$a_{v\cdot} = \sum_{i=1}^{n}\sum_{j=1}^{m_i} I_{V=v} y_{ij}, \qquad N_{v\cdot} = \sum_{i=1}^{n}\sum_{j=1}^{m_i} I_{V=v}.$$

Let SAR_{vs} denote the secondary attack rate from an index case with vaccine status v to a susceptible with vaccine status s. Pooling across compounds, the two standard SARs not stratified by the vaccine status of the index case used in estimating protective VE_S (Orenstein et al. 1988) are $SAR_{\cdot s} = a_{\cdot s}/N_{\cdot s}$, $s = 0, 3$. If not stratified by vaccine status of the susceptible, $SAR_{v\cdot} = a_{v\cdot}/N_{v\cdot}$, $v = 0, 3$. The nonparametric estimates of the four SARs stratified by vaccine status of index cases and susceptibles are $SAR_{vs} = a_{vs}/N_{vs}$, $v, s = 0, 3$.

3.4 Data

A total of 518 of the 1,800 compounds (29%) were detected as having potential cases of pertussis in 1993. Pertussis was confirmed in 189 (36%) of those compounds, representing 232 primary and coprimary cases and 1,217 susceptibles. Among these, compounds with coprimary cases [$n = 33$ (17%)], compounds with no susceptibles [$n = 5$ (3%)], and compounds with a partially vaccinated primary case [$n = 42$ (22%)] were excluded. Thus 109 of the 189 of the qualifying compounds (58%) were eligible for analysis. They represented 109 primary cases and 790 susceptibles, of whom 152 (19%) were partially vaccinated and 638 (81%) were either unvaccinated or completely vaccinated. Only the latter group of 638 was considered in the analysis presented here. The effect of partial vaccination has been discussed elsewhere (Préziosi and Halloran 2003). The result of at least one biological confirmation criterion was available in more than 97% of the suspected cases meeting the clinical definition used in this study.

Table 1 gives the data from Niakhar, along with the nonparametric SAR estimates for this case definition and the four different cutoffs. The number of exposed susceptibles per compound ranged from 1 to 32 (interquartile range, 2–8), with a median of 5 and a mean of 5.85. Using no cutoff, there were 154 secondary cases in 638 susceptibles exposed to the 109 unique index cases, for an overall SAR of .24. This article presents VE results using no cutoff; the full analysis included estimating VE using the four different cutoffs.

4. ESTIMATING VACCINE EFFICACY

4.1 Vaccine Efficacy Based on Nonparametric Secondary Attack Rate

The conventional pooled, nonparametric estimator of protective VE, the effect of vaccination on susceptibility, VE_S, not stratified by vaccine status of the index cases, is (Orenstein et al. 1988)

$$VE_{S\cdot 3/\cdot 0} = 1 - \frac{SAR_{\cdot 3}}{SAR_{\cdot 0}}. \tag{1}$$

The analogous pooled estimator of VE_I, not stratified by the vaccination status of the susceptibles, and the vaccine effect if both index case and susceptible are vaccinated, VE_T, are (Halloran et al. 1997)

$$VE_{I3\cdot/0\cdot} = 1 - \frac{SAR_{3\cdot}}{SAR_{0\cdot}}, \qquad VE_T = 1 - \frac{SAR_{33}}{SAR_{00}}. \tag{2}$$

These are the three main vaccine effects of interest. However, the protective effect of vaccination might be different if the index case were vaccinated than if he or she were unvaccinated. Similarly, the effect of vaccination of the index cases on transmission might be different in susceptibles who are vaccinated than those who are unvaccinated; that is, the vaccine status in either the index case or the susceptible could modify the VE. The stratified measures of VE_S and VE_I are

$$\begin{aligned} VE_{S03/00} &= 1 - \frac{SAR_{03}}{SAR_{00}}, & VE_{S33/30} &= 1 - \frac{SAR_{33}}{SAR_{30}}, \\ VE_{I30/00} &= 1 - \frac{SAR_{30}}{SAR_{00}}, & VE_{I33/03} &= 1 - \frac{SAR_{33}}{SAR_{03}}. \end{aligned} \tag{3}$$

A commonly used confidence interval (CI) is obtained from the expression for the confidence interval of log relative risk (Katz et al. 1978). For example, from Table 1, $SAR_{30} = 9/67 = .13$ and $SAR_{00} = 73/198 = .37$, so $VE_{I30/00} = 1 - (9/67)/(73/198) = .64$, 95% CI [.31, .81]. However, neither the simple pooled SAR estimator nor the confidence interval takes into account the structure of the correlated binary data.

Table 1. Number of Exposed Susceptibles, Secondary Cases, and SAR by Vaccination Status of the Index Case and the Exposed Susceptible Children and Cutoff for Counting Secondary Cases

	Exposed susceptibles and secondary cases					
	Vaccinated		*Unvaccinated*		*Combined*	
Index case	*Cases/exposed*	*SAR*	*Cases/exposed*	*SAR*	*Cases/exposed*	*SAR*
Vaccinated						
cutoff: none	11/127	.09	9/67	.13	20/194	.10
56 days	10/127	.08	6/67	.09	16/194	.08
42 days	10/127	.08	5/67	.07	15/194	.08
28 days	3/127	.02	3/67	.04	6/194	.03
Unvaccinated						
cutoff: none	61/246	.25	73/198	.37	134/444	.30
56 days	55/246	.22	67/198	.34	122/444	.27
42 days	52/246	.21	66/198	.33	118/444	.27
28 days	41/246	.17	52/198	.26	93/444	.21
Combined						
cutoff: none	72/373	.19	82/265	.31	154/638	.24
56 days	65/373	.17	73/265	.28	138/638	.22
42 days	62/373	.17	71/265	.27	133/638	.21
28 days	44/373	.11	55/265	.21	99/638	.16

4.2 Vaccine Efficacy Based on the Logistic Model

To take correlation within compounds into account, we consider a marginal model and a random-effects model. The parametric form in both cases is the logistic model, with the SAR as the usual probability, p, of an event. Another advantage of the model-based approach is that additional covariates can be easily included, for example, age and gender either of the index case as compound-level environmental variables or of the susceptibles as individual variables. In the following comparison, we borrow heavily from the excellent exposition of Diggle, Liang, and Zeger (1994, pp. 131–135). In marginal models, inference about population averages is the focus. If there is heterogeneity across compounds in the baseline transmission, then the estimated baseline coefficients represent an average over the heterogeneities. The correlation structure is some function of the marginal mean and possibly additional parameters. In the random-effects model, a slightly different baseline transmission is estimated for each compound, with the degree of heterogeneity estimated in the variance of the random effect. The vaccine effects in each compound are interpreted in relation to that compound's baseline transmission. In this application, our primary scientific question is about the population average, or marginal, VE measures, so the marginal model is our model of choice. Nonetheless, we present examples of both marginal and random-effects models to illustrate differences in interpretation and results. The coefficients for the marginal and random-effects models are indicated by $\boldsymbol{\beta}$ and $\boldsymbol{\beta}^*$.

4.3 The Marginal Model

The marginal model for the logit of the SAR_{ij} of the jth person in the ith household is

$$\operatorname{logit}(SAR_{ij}) = \beta_0 + \beta_1 x_{i\cdot 1} + \beta_2 x_{ij2}, \tag{4}$$

where $x_{i\cdot 1}$ denotes the vaccine status of the index case in compound i and x_{ij2} denotes the vaccine status of the jth exposed susceptible in compound i. The vaccine status of the index case, $x_{i\cdot 1}$, enters the analysis as a compound-level environmental variable. Because we are interested in VE estimates on the SAR scale, we transform the parameters from the logistic model to the probability scale. The stratified SARs from model (4) are

$$\begin{aligned} SAR_{00} &= \frac{\exp\beta_0}{1+\exp\beta_0}, \quad SAR_{03} = \frac{\exp(\beta_0+\beta_2)}{1+\exp(\beta_0+\beta_2)}, \\ SAR_{30} &= \frac{\exp(\beta_0+\beta_1)}{1+\exp(\beta_0+\beta_1)}, \quad SAR_{33} = \frac{\exp(\beta_0+\beta_1+\beta_2)}{1+\exp(\beta_0+\beta_1+\beta_2)}. \end{aligned} \tag{5}$$

Parameter estimates from the foregoing model provide estimates for the stratified $\mathrm{VE}_{S00/03}$ and $\mathrm{VE}_{S30/33}$, the stratified $\mathrm{VE}_{I00/30}$ and $\mathrm{VE}_{I03/33}$, and VE_T. Plugging the expressions for the SARs into (1) and (2), the expressions for the VE measures are

$$\begin{aligned} VE_{S03/00} &= \frac{1-\exp(\beta_2)}{1+\exp(\beta_0+\beta_2)}, \\ VE_{S33/30} &= \frac{1-\exp(\beta_2)}{1+\exp(\beta_0+\beta_1+\beta_2)}, \\ VE_{I30/00} &= \frac{1-\exp(\beta_1)}{1+\exp(\beta_0+\beta_1)}, \\ VE_{I33/03} &= \frac{1-\exp(\beta_1)}{1+\exp(\beta_0+\beta_1+\beta_2)}, \\ VE_T &= \frac{1-\exp(\beta_1+\beta_2)}{1+\exp(\beta_0+\beta_1+\beta_2)}. \end{aligned} \tag{6}$$

To obtain estimates of the unstratified $\mathrm{VE}_{I3./0.}$ and $\mathrm{VE}_{S.3/.0}$, we can fit additional submodels, such as $\operatorname{logit}(SAR_{ij}) = \beta'_0 + \beta'_1 x_{i\cdot 1}$ and $\operatorname{logit}(SAR_{ij}) = \beta''_0 + \beta''_2 x_{ij2}$, and transform back to get

$$\begin{aligned} VE_{I3./0.} &= \frac{1-\exp(\beta'_1)}{1+\exp(\beta'_0+\beta'_1)}, \\ VE_{S.3/.0} &= \frac{1-\exp(\beta''_2)}{1+\exp(\beta''_0+\beta''_2)}. \end{aligned} \tag{7}$$

Alternatively, we could use the parameter estimates from the full model (4) and substitute the respective means of $x_{i\cdot 1}$ and x_{ij2}. We present results for $\mathrm{VE}_{I3./0.}$ and $\mathrm{VE}_{S.3/.0}$ for the two submodels.

We estimated the marginal model taking into account correlation of transmission within compound using generalized estimating equations (GEEs) (Liang and Zeger 1986). The analysis was done using the "repeated" option in PROC GENMOD in SAS version 8.2 (SAS Institute, Inc., 1999), assuming an exchangeable working correlation matrix.

We obtained appropriate confidence intervals on the transformed scale using the bootstrap (Efron and Tibshirani 1993). Bootstrap samples were selected using the compound as the sampling unit. We estimated the GEE logistic regression coefficients for each bootstrap sample, then transformed them to the probability scale to get the VE estimates for that bootstrap sample. Three different bootstrap confidence intervals were computed: the percentile, the bias-corrected (BC), and the bias-corrected and accelerated (BC_a) intervals. Confidence intervals were based on 2,000 bootstrap samples (Efron and Tibshirani 1993, p. 275). Each of the three types of bootstrap CIs are based on the same collection of 2,000 bootstrap estimates, just the upper and lower limits are determined differently. The bias correction is derived simply from the proportion of the bootstrap estimates lower than the point estimate of the data. Estimating the acceleration constant requires jackknifing the data. Bootstrap confidence intervals sampling on compounds were also computed for the VE estimators based on the nonparametric SARs described in the previous section (Kolczak and Halloran 1995). We computed analytic confidence intervals for the GEE estimates of VE on the transformed scale using the multivariate delta method (Agresti 1990; Dunson and Halloran 1996).

4.4 The Random-Effects Model

The random-effects model for the logit of the SAR_{ij} of the jth person in the ith household is $\operatorname{logit}(SAR_{ij}|U_i) = (\beta^*_0 + U_i) + \beta^*_1 x_{i\cdot 1} + \beta^*_2 x_{ij2}$. The simplest model assumes the random effect $U_i \sim N(0, \sigma^2)$. If we were to remain on the logistic scale, then the parameter β^*_0 would be interpreted as the log-odds of transmission from an unvaccinated index case to an unvaccinated susceptible for a typical compound with random

effect $U_i = 0$. The parameter β_1^* would be the log-odds ratio of transmission occurring when the index case is vaccinated compared to when it is unvaccinated within any given compound. The parameter β_2^* would be the log-odds ratio of transmission occurring when a susceptible in the compound is vaccinated compared with a susceptible in that same compound who is unvaccinated.

But we are interested in transforming to the SAR scale to obtain the different VE estimates. We obtained the compound-specific SAR_{ij}s by incorporating the random effect into the expression. For example, the compound-specific SAR_{00i} from an unvaccinated index case to an unvaccinated susceptible is $SAR_{00i}|U_i = \exp(\beta_0^* + U_i)/[1 + \exp(\beta_0^* + U_i)]$. The marginal SAR_{00} is the estimated expectation of the SAR_{00} obtained by numerical integration over the estimated distribution of the random effects. The VE_i estimates for each compound are obtained from expressions analogous to (6). The marginal VE estimates are the estimated expectations obtained by numerical integration over the estimated distribution of the random effects. To obtain estimates of the unstratified $VE_{I3./0.}$ and $VE_{S.3/.0}$, we fit random-effects submodels similar to those described earlier.

We used two methods to estimate the random-effects model. The first is a Bayesian hierarchical model (Carlin and Louis 2000); the second is a nonlinear mixed model (Davidian and Giltinan 1995). In the Bayesian hierarchical model, the compound specific random effects were assumed to be $U_i \sim N(0, 1/\tau)$, with a hyperprior of $\tau \sim gamma(.0001, .0001)$. Vague priors of $N(0, 10^6)$ were assumed for the β_i^*s, $i = 0, 1, 2$. Computation was done using Markov chain Monte Carlo (MCMC) methods in WinBUGS (Spiegelhalter, Thomas, and Best 2000). Burn-in consisted of 1,000 iterations, with 5,000 iterations used for posterior summaries. Convergence was assessed using the Gelman and Rubin convergence statistic (Gelman and Rubin 1992; Brooks and Gelman 1998). For each iteration, we computed the SAR and VE measures for each compound. The population mean VE measures were computed by averaging over the compounds at each iteration. The 95% posterior credible intervals for the VE measures are available directly on the transformed scale from the approximation to the posterior distribution from the MCMC chains.

The nonlinear mixed model was fit using PROC NLMIXED Wolfinger 1999) in SAS version 8.2 (SAS Institute Inc., 1999). The U_i were assumed to be iid $N(0, \sigma^2)$. PROC NLMIXED maximizes an approximation to the likelihood integrated over the random effects directly (Pinheiro and Bates 1995) and produces empirical Bayes estimates of the random effects (Wolfinger 1999). We used the adaptive Gaussian quadrature approximation and dual quasi-Newton algorithm optimization techniques in PROC NLMIXED. Confidence intervals on the transformed SAR scale were obtained using the bootstrap as described earlier. The marginal VE estimates were calculated for each bootstrap sample using numerical integration over the estimated random-effects distribution.

5. RESULTS

5.1 Baseline Secondary Attack Rates

Estimates of the baseline SAR_{00} for each model are shown in Figure 1. Horizontal lines represent point estimates of the nonparametric pooled SAR_{00}, the marginal GEE model, and the estimated expected SAR_{00} on the transformed scale for the

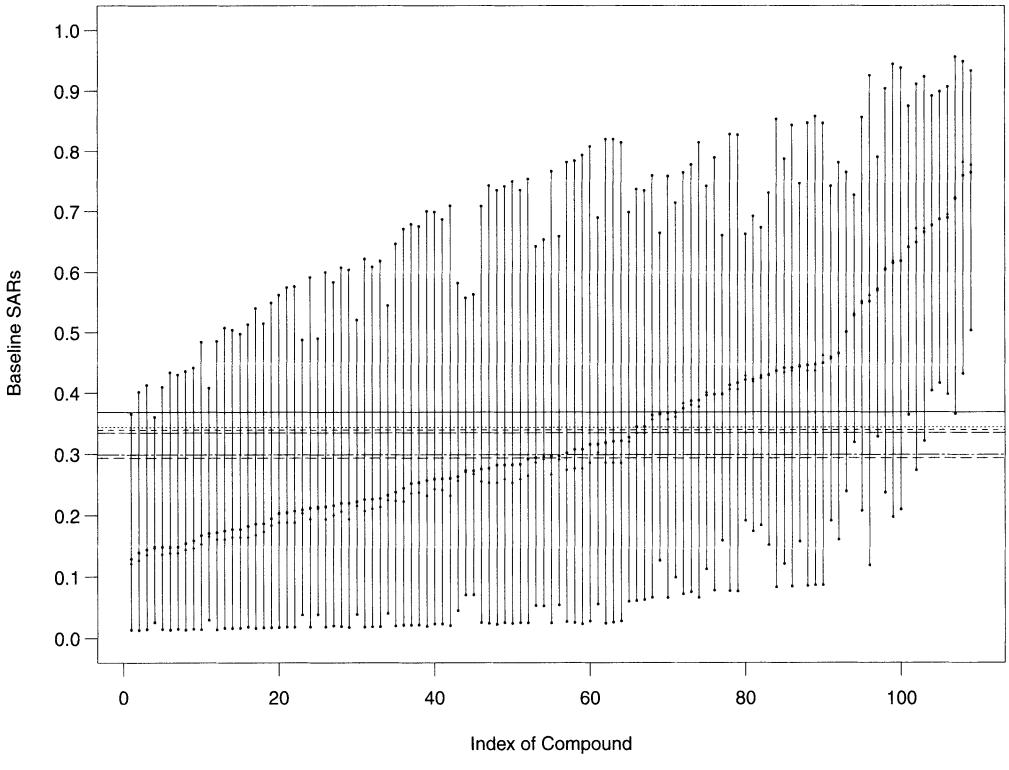

Figure 1. Baseline SAR_{00} for Each Model, Represented by Horizontal Lines. Compound-level baseline SAR_{00i}'s for the random-effects nonlinear mixed model (NLMIXED) are also plotted. The posterior mean and 95% posterior CI are plotted for each Bayesian compound-level SAR_{00i}. The indexes of the compounds are in order of increasing mean from the MCMC chain. ——— nonparametric; ····· GEE; – – – – Gibbs sampler, mean; — — — Gibbs sampler; ——— NLMIXED, mean; ——-——- NLMIXED; • Gibbs sampler; • NLMIXED.

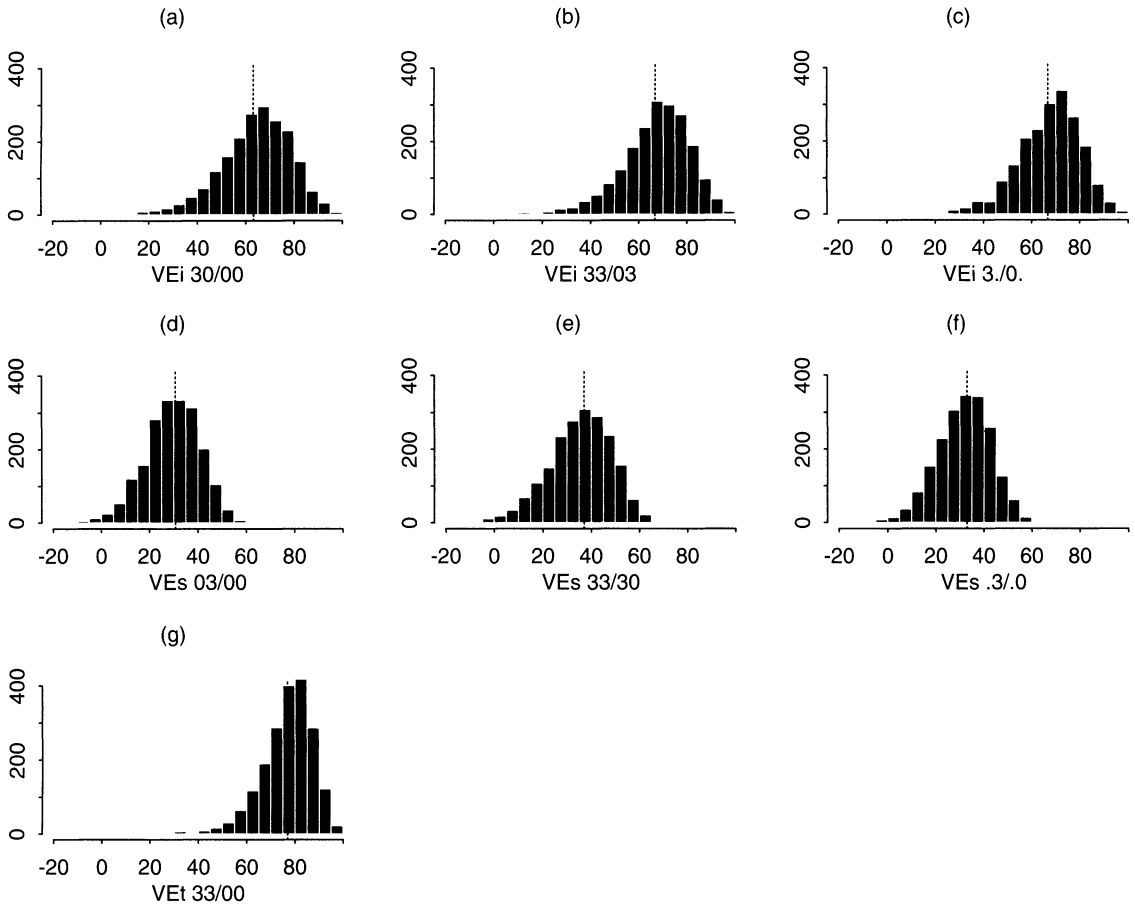

Figure 2. Histograms of 2,000 Bootstrap Estimates of (a–c) VE for Infectiousness, VE_I, Stratified and Unstratified; (d–f) VE for Susceptibility; VE_S, Stratified and Unstratified; and (g) Total VE, VE_T, Based on the GEE Logistic Regression Parameters. The dotted line in each histogram indicates the estimate for the actual dataset.

two random-effects models. Included for comparison are horizontal lines for the estimate of the nonlinear mixed model and the mean from the Bayesian MCMC chain of the SAR_{00i} when $U_i = 0$. The nonparametric estimate of SAR_{00} is the highest. The estimated expectations of the conditional baselines for the random-effects models are similar to the baseline SAR_{00} from the marginal model.

Figure 1 shows the SAR_{00i}s for each of the 109 compounds for the random-effects models, the point estimates of the SAR_{00i} for the nonlinear mixed model, and the mean SAR_{00i} and 95% posterior CI from the MCMC chain for the Bayesian model. The individual estimated SAR_{00i}s range from about .1 to .7, although 51 of the 109 compounds had no secondary cases. The estimated variance of the random effect is 1.210^2 for the full NLMIXED model and 1.272^2 for the Bayesian model. The distribution of the U_i estimates is slightly right-skewed. The upper 17 SAR_{00i}s increase more rapidly than the others, suggesting a slight departure from normality. No explanatory variables were found to be associated with the increase. If one wanted to pursue the random-effects approach, one might consider modeling the data as arising from two populations with different random-effects distributions.

5.2 Vaccine Effects

Figure 2 shows the point estimates and histograms of 2,000 bootstrap estimates of the VE_I, VE_S, and VE_T parameters based on the GEE model. The higher values of the VE_I and VE_T produce more skewed histograms of estimates than does the VE_S. Although there is little evidence of effect modification in the stratified VE_I and VE_S estimates, we present them for completeness. Because all of the VE_I bootstrap estimates are well above 0, the lower limits of the 95% CIs will also be well above 0.

Figure 3 shows the different point estimates and confidence intervals for VE_S, VE_I, and VE_T. Table 2 contains selected results. We present all three bootstrap CIs only for the GEE model to illustrate the behavior at the different efficacies. The qualitative behavior is the same for all models for which we did the bootstrap. To simplify the presentation, we show just the BC interval in most cases, because it lies between the percentile and the BC_a intervals (Fig. 3).

The point estimates for VE_I and VE_T obtained from the nonparametric SAR and from the GEE are nearly identical. The estimates from the two random-effects models, particularly for VE_I and VE_T, are higher than either the GEE or SAR estimates.

The bootstrap CIs for the nonparametric VE estimates are wider than the simple CIs based on the log-relative risk. In particular, the bootstrap CIs for VE_I and, to a lesser extent, VE_T are wider. For example, the BC bootstrap 95% CI of $VE_{I3./0.}$ is 1.94 wider than the simple 95% CI. The difference is less pronounced with CIs of VE_S, with the ratio of the lengths being between 1.2 and 1.3. Thus the conventionally used CI substantially underrepresents the variability in the data. The greater

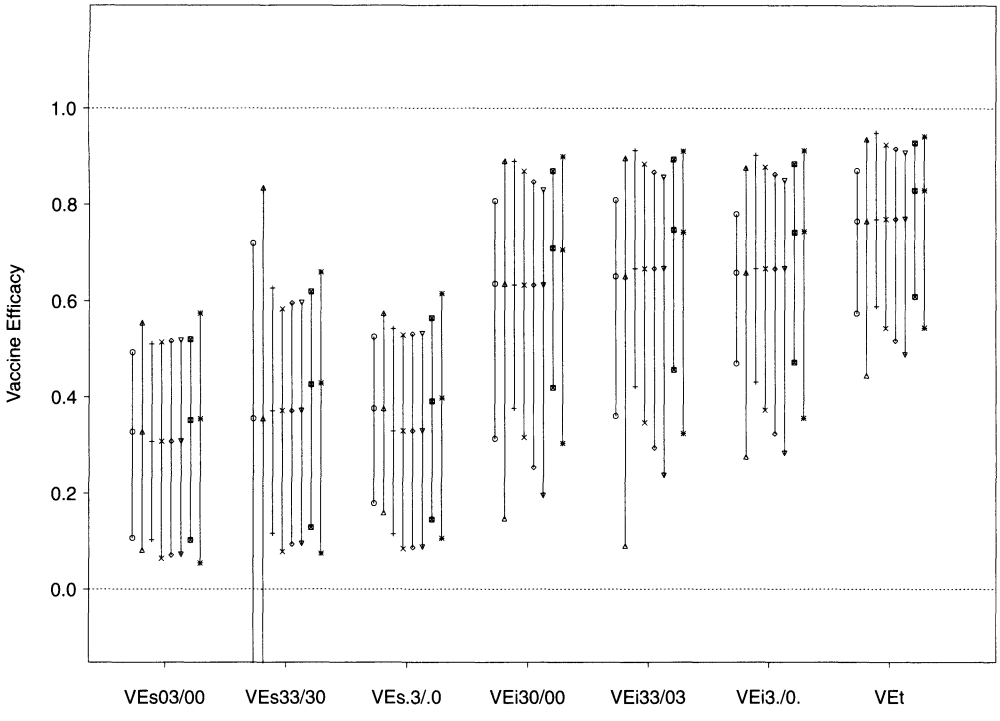

Figure 3. Comparison of Different Estimators and Confidence Intervals for VE for Susceptibility, VE_S, stratified and unstratified; VE for infectiousness, VE_I, stratified and unstratified; and total VE, VE_T. o—o—o SAR, binomial distribution; △—△—△ SAR, BC CI; +—+—+ GEE, multivariate delta method; ×—×—× GEE, percentile CI; ◇—◇—◇ GEE, BC CI; ▽—▽—▽ GEE, BCa CI; □—□—□ Bayesian hierarchical model; ∗—∗—∗ NLMIXED, BC CI.

sensitivity of the variability of the VE_I and VE_T estimators to compound-level effects might result from the vaccine status of the index case being a compound-level environmental variable. The nonparametric estimate of $VE_{S33/30}$ is unstable because the total number of secondary cases was only 20, compared with 134 cases for $VE_{S03/00}$, so both the simple and the BC bootstrap CIs are quite wide.

The bootstrap CIs of the GEE estimates of VE_I are also wider than those based on the simple CI for the nonparametric VE estimates, but not as wide as the bootstrap CIs of the nonparametric VE estimates. For example, the GEE percentile, BC, and BC_a bootstrap 95% CIs for $VE_{I3./0.}$ are 1.63, 1.74, and 1.83 wider than the simple SAR 95% CI. Thus the parametric model in the GEE helps stabilize the estimation compared with the nonparametric approach.

At the higher VE_I and VE_T, the three bootstrap CIs for the GEE estimates differ more from one another than at the lower VE_S, due to the skewness at higher efficacies. The consistent decrease of the lower CI limits moving from percentile, to BC, and to BC_a for VE_I and VE_T result from the bias correction and the acceleration constant moving in the same direction; thus the BC_a intervals are well behaved. The BC_a intervals for the other estimators were also well behaved.

The multivariate delta method CIs on the GEE estimates are symmetric and similar in length to the percentile bootstrap CIs. However, the normality assumption of the VE_I and VE_T estimators is clearly violated, so we do not recommend using the multivariate delta method. Also, CIs based on the multivariate delta method could theoretically exceed 1, which could cause difficulty because VE is bounded at 1.

6. DISCUSSION

Because the scientific questions of interest are the population average vaccine effects, the GEE approach to estimate the marginal effects of vaccination while taking into account the clustered binary structure of the data is appropriate. The GEE

Table 2. Pertussis VE Estimates From the Niakhar Region, Senegal, 1993

	VE × 100% (95% confidence interval)						
	VE for susceptibility			VE for infectiousness			
Estimator	$VE_{S03/00}$	$VE_{S33/30}$	$VE_{S.3/.0}$	$VE_{I30/00}$	$VE_{I33/03}$	$VE_{I3./0.}$	VE_T
SAR (BC*)	$33_{(8,55)}$	$36_{(-62,88)}$	$38_{(16,57)}$	$64_{(15,89)}$	$65_{(9,90)}$	$66_{(28,88)}$	$77_{(45,94)}$
SAR (simple)	$33_{(11,49)}$	$36_{(-48,72)}$	$38_{(18,53)}$	$64_{(31,81)}$	$65_{(36,81)}$	$66_{(47,78)}$	$77_{(58,87)}$
GEE (BC)	$31_{(7,52)}$	$37_{(9,60)}$	$33_{(9,53)}$	$63_{(25,85)}$	$67_{(29,87)}$	$67_{(32,86)}$	$77_{(52,92)}$
NLMIXED (BC)	$35_{(5,57)}$	$43_{(7,66)}$	$40_{(11,61)}$	$71_{(32,90)}$	$74_{(32,91)}$	$74_{(36,91)}$	$83_{(54,94)}$
Bayes median	$35_{(10,52)}$	$43_{(13,62)}$	$39_{(15,56)}$	$71_{(42,87)}$	$75_{(46,89)}$	$74_{(47,88)}$	$83_{(61,93)}$

*BC, bias-corrected bootstrap confidence interval.

approach also produces estimates similar to the conventional nonparametric estimators. Besides the usual measure of protective efficacy, VE_S, we have demonstrated estimation of VE_I and VE_T from SARs. In particular, when estimating VE_I, a relatively new aspect of evaluating vaccines, the CIs were considerably wider when the correlated data structure was taken into account. If by design, each transmission unit has only one exposed susceptible, then the assumption of independence of susceptibles is valid, and the pooled nonparametric SAR approach with log-relative risk CI is appropriate.

The nonparametric pooled SAR log-relative risk confidence intervals are easy to use for an exploratory analysis of the data. However, the GEE estimates are computed easily in SAS from PROC GENMOD. The main trick is in obtaining the bootstrap confidence intervals. The compounds in each bootstrap sample need to be renumbered so that PROC GENMOD will think that each compound is unique, even though it may be in the sample more than once. The bootstrap can also be used with the pooled nonparametric SAR approach. If there are sufficient data, then the bootstrap CIs are similar for both the GEE parametrics and the nonparametric estimates. However, with sparse data, the parametric model produces substantially narrower intervals.

Our focus here has been on appropriate inference within the limits of the conventional SAR analysis (Fine et al. 1988; Orenstein et al. 1988), which makes several assumptions (Halloran 2001). They include (1) the transmission units are independent, (2) the incubation and latent periods are fixed, (3) the infectious period is fixed, and (4) coprimaries are irrelevant.

Assuming independent transmission units implies the asymmetric assumption that the index case and coprimaries get infected from outside the unit, while the susceptibles are exposed only within the unit. Other models that do not assume independence (Longini and Koopman 1982; Longini, Koopman, Haber, and Cotsonis 1988; Rampey, Longini, Haber, and Monto 1992; Longini et al. 1996) require more assumptions about the community mixing structure that may not be valid. Although SAR estimates are sensitive to the mixing assumptions, ratio estimators such as VE are more robust to changes in model specification (Longini Halloran, Haber, and Chen 1993).

Because times of infection and onset and termination of infectiousness are not observed, estimation of the distribution of the latent, incubation, and infectious periods is a challenge (Bailey 1975; O'Neill, Balding, Becker, Eerola, and Mollison 2001). The conventional SAR analysis makes use of information from other sources to determine the lengths of the periods (Kendrick and Eldering 1939). We recommend exploring different cutoff times as a sensitivity analysis of how the SARs and vaccine effects vary with the assumption of the infectious period. The VE ratio estimators are less sensitive to the choice of cutoff then the SARs. The VE_I estimates tended to be higher with a shorter cutoff time. For example, at 28 days cutoff, the GEE estimate of $VE_{I3./0.} = 85$, 95% BC_a CI [46, 95] (Préziosi and Halloran 2003), higher than that presented here for no cutoff. Estimates based on no cutoff are likely the more conservative. Later cases would be less likely to have been infected by the index case, so they dilute the effect estimates.

Ignoring coprimaries in the SAR analysis carries with it the strong assumption that the probability of being infected is independent of the number of infectious contacts, so we do not recommend this. Here we removed compounds that had coprimaries. Chu, Préziosi, and Halloran (2001) developed a method to include transmission units with coprimaries with differing vaccine status. Reanalysis of these data including compounds with coprimaries changed neither the point estimates nor the confidence intervals substantially.

Here we emphasized estimation of VE_I and VE_T, relatively new measures for evaluating vaccination. The properties of VE_I require further research. The people who become infected and expose other people are not a random sample of the population; thus there is a potential for selection bias (Halloran and Struchiner 1995). Also, studies designed to estimate VE_I will need larger samples sizes than studies to estimate VE_S.

Does the analysis provide convincing evidence that pertussis vaccination reduces transmission of *Bp*? The outcomes are based on clinical disease, not infection. The results are most consistent with the interpretation that in fact vaccination reduces transmission of infection. Further answers to questions of how pertussis vaccination affects transmission require studies designed to look at infection, rather than disease. In this article we have presented methods to improve estimation of VE from the conventional SARs. Much challenging statistical research remains to be done.

[Received November 2001. Revised November 2002.]

REFERENCES

Aaby, P., Samb, B., Anderson, M., and Simondon, F. (1996), "No Long-Term Excess Mortality After Measles Infection: A Community Study From Senegal," *American Journal of Epidemiology*, 143, 1035–1041.

Agresti, A. (1990), *Categorical Data Analysis*, New York: Wiley.

Bailey, N. T. J. (1975), *The Mathematical Theory of Infectious Diseases and Its Application* (2nd ed.), London: Griffin.

Brooks, S. P., and Gelman, A. (1998), "Alternative Methods for Monitoring Convergence of Iterative Simulations," *Journal of Computational Graphics and Statistics*, 7, 434–455.

Carlin, B. P., and Louis, T. A. (2000), *Bayes and Empirical Bayes Methods for Data Analysis* (2nd ed.), New York: Chapman and Hall/CRC.

Chu, H., Halloran, M. E., and Préziosi, M.-P. (2001), "Estimating Heterogeneous Transmission With Multiple Primary Cases Using MCMC Methods," Technical Report 01-11, Emory University, Dept. of Biostatistics.

Davidian, M., and Giltinan, D. M. (1995), *Nonlinear Models for Repeated Measurement Data*, New York: Chapman and Hall/CRC.

Diggle, P. J., Liang, K. Y., and Zeger, S. L. (1994), *Analysis of Longitudinal Data*, Oxford, UK: Oxford University Press.

Dunson, D. B., and Halloran, M. E. (1996), "Estimating Transmission Blocking Efficacy of Malaria Vaccines," Technical Report 96-16, Emory University, Dept. of Biostatistics.

Efron, B., and Tibshirani, R. J. (1993), *An Introduction to the Bootstrap*, New York: Chapman and Hall.

Fine, P. E. M., and Clarkson, J. A. (1982), "The Recurrence of Whooping Cough: Possible Implications for Assessment of Vaccine Efficacy," *Lancet*, 1, 666–669.

——— (1987), "Reflections on the Efficacy of Pertussis Vaccines," *Reviews of Infectious Diseases*, 9, 866–883.

Fine, P. E. M., Clarkson, J. A., and Miller, E. (1988), "The Efficacy of Pertussis Vaccines Under Conditions of Household Exposure: Further Analysis of the 1978–80 PHLS-ERL Study in 21 Area Health Authorities in England," *International Journal of Epidemiology*, 17, 635–642.

Francis, T., Korns, R. F., Voights, R. B., Boisen, M., Hemphill, F. M., Napier, J. A., and Tolchinsky, E. (1955), "An Evaluation of the 1954 Poliomyelitis Vaccine Trials," *American Journal of Public Health*, 45, 1–63.

Garenne, M., and Cantrelle, P. (1998), "Three Decades of Research on Population and Health: The Orstom Experience in Rural Senegal, 1962–1991,"

in *Prospective Community Studies in Developing Countries*, eds. M. Das Gupta, P. Aaby, M. Garenne, and G. Pison; Oxford, UK: Clarendon Press, pp. 233–252.

Garenne, M., Leroy, O., Beau, J.-P., and Sene, I. (1993), "Efficacy of Measles Vaccines After Controlling for Exposure," *American Journal of Epidemiology*, 138, 182–195.

Gelman, A., and Rubin, D. B. (1992), "Inference From Iterative Simulation Using Multiple Sequences," *Statistical Sciences*, 7, 457–472.

Golm, G. T., Halloran, M. E., and Longini, I. M. (1998), "Semiparametric Models for Mismeasured Exposure Information in Vaccine Trials," *Statistics in Medicine*, 17, 2335–2352.

Halloran, M. E. (2001), "Overview of Study Design," in *Epidemiologic Methods for the Study of Infectious Diseases*, eds. J. C. Thomas and D. J. Weber, Oxford, UK: Oxford University Press.

Halloran, M. E., and Struchiner, C. J. (1995), "Causal Inference for Infectious Diseases," *Epidemiology*, 6, 142–151.

Halloran, M. E., Struchiner, C. J., and Longini, I. M. (1997), "Study Designs for Different Efficacy and Effectiveness Aspects of Vaccination," *American Journal of Epidemiology*, 146, 789–803.

Katz, D., Baptista, J., Azen, S. P., and Pike, M. C. (1978), "Obtaining Confidence Intervals for the Risk Ratio in Cohort Studies," *Biometrics*, 34, 469–474.

Kendrick, P., and Eldering, G. (1939), "A Study in Active Immunization Against Pertussis," *American Journal of Hygiene, Section B*, 38, 133.

Kim-Farley, R., Bart, S., Stetler, H., Orenstein, W., Bart, K., Sullivan, K., Halpin, T., and Sirotkin, B. (1985), "Clinical Mumps Vaccine Efficacy," *American Journal of Epidemiology*, 121, 593–597.

Kolzcak, M., and Halloran, M. E. (1995), "Estimating the Variability of Secondary Attack Rates and Vaccine Efficacy Based on Secondary Attack Rates," Technical Report 95-7, Emory University, Dept. of Biostatistics.

Liang, K. Y., and Zeger, S. L. (1986), "Longitudinal Data Analysis Using Generalized Linear Models," *Biometrika*, 73, 13–22.

Longini, I. M., Datta, S., and Halloran, M. E. (1996), "Measuring Vaccine Efficacy for Both Susceptibility to Infection and Reduction in Infectiousness for Prophylactic HIV-1 Vaccines," *Journal of Acquired Immune Deficiency Syndrome and Human Retrovirus*, 13, 440–447.

Longini, I. M., Halloran, M. E., Haber, M. J., and Chen, R. T. (1993), "Measuring Vaccine Efficacy From Epidemics of Acute Infectious Agents," *Statistics in Medicine*, 12, 249–263.

Longini, I. M., and Koopman, J. S. (1982), "Household and Community Transmission Parameters From Final Distributions of Infections in Households," *Biometrics*, 38, 115–126.

Longini, I. M., Koopman, J. S., Haber, M. J., and Cotsonis, G. A. (1988), "Statistical Inference for Infectious Diseases: Risk-Specified Household and Community Transmission Parameters," *American Journal of Epidemiology*, 128, 845–859.

Miller, E., and Gay, N. J. (1997), "Epidemiological Determinants of Pertussis," *Developments in Biological Standardization*, 89, 15–23.

O'Neill, P. O. 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.

Orenstein, W. A., Bernier, R. H., Dondero, T. J., Hinman, A. R., Marks, J. S., Bart, K. J. and Sirotkin, B. (1985), "Field Evaluation of Vaccine Efficacy," *Bulletin of the World Health Organization*, 63, 1055–1068.

Orenstein, W. A., Bernier, R. H., and Hinman, A. R. (1988), "Assessing Vaccine Efficacy in the Field: Further Observations," *Epidemiologic Reviews*, 10, 212–241.

Pinheiro, J. C., and Bates, D. M. (1995), "Approximations to the Log-Likelihood Function in the Nonlinear Mixed-Effects Model," *Journal of Computation and Graphical Statistics*, 4, 12–35.

Préziosi, M.-P., and Halloran, M. E. (2003), "Effects of Pertussis Vaccination on Transmission: Vaccine Efficacy for Infectiousness," *Vaccine*, in press.

Préziosi, M.-P., Yam, A., Ndiaye, M., Simaga, A., Simondon, F., and Wassilak, S. G. (1997), "Practical Experiences in Obtaining Informed Consent for a Vaccine Trial in Rural Africa," *New England Journal of Medicine*, 336, 370–373.

Préziosi, M.-P., Yam, A., Wassilak, S. G., Chabirand, L., Simaga, A., Ndiaye, M., Dia, F., Dabis, F., and Simondon, F. (2002), "Epidemiology of Whooping Cough in a West African Community Before and After Introduction of a Widespread Vaccination Programme," *American Journal of Epidemiology*, 155, 891–896.

Rampey, A. H., Longini, I. M., Haber, M. J., and Monto, A. S. (1992), "A Discrete-Time Model for the Statistical Analysis of Infectious Disease Incidence Data," *Biometrics*, 48, 117–128.

Rohani, P., Earn, D. J., and Grenfell, B. T. (2000), "Impact of Immunisation on Pertussis Transmission in England and Wales" [letter], *Lancet*, 355, 285–286.

SAS Institute, Inc. (1999), *SAS Software Version 8.1 for the Unix Operating System (SunOS)*, Cary, NC: author.

Simondon, F., Iteman, I., Préziosi, M.-P., Yam, A., and Guiso, N. (1998), "Evaluation of an Immunoglobulin G Enzyme-Linked Immunosorbent Assay for Pertussis Toxin and Filamentous Hemagglutinin in Diagnosis of Pertussis in Senegal," *Clinical and Diagnostic Laboratory Immunology*, 5, 130–134.

Simondon, F., Préziosi, M.-P., Yam, A., Kane, C. T., Chabirand, L., Iteman, I., Sanden, G., Mboup, S., Hoffenbach, A., Knudsen, K., Guiso, N., Wassilak, S., and Cadoz, M. (1997), "A Randomized Double-Blind Trial Comparing a Two-Component Acellular to a Whole-Cell Pertussis Vaccine in Senegal," *Vaccine*, 15, 1606–1612.

Spiegelhalter, D. J., Thomas, A., and Best, N. G. (2000), WinBUGS User Manual, version 1.3. Medical Research Council Biostatistics Unit, Cambridge, UK.

Storsaeter, J., Blackwelder, W. C., and Hallander, H. O. (1992), "Pertussis Antibodies, Protection, and Vaccine Efficacy After Household Exposure," *American Journal of Diseases of Children*, 146, 167–172.

Taranger, J., Trollfors, B., Bergfors, E., Knutsson, N., Sundh, V., Lagergard, T, Lind-Brandberg, L., Zackrisson, G., White, J., Cicierllo, H., Fusco, J., and Robbins, J. B. (2001), "Mass Vaccination of Children With Pertussis Toxoid—Decreased Incidence in Both Vaccinated and Nonvaccinated Persons," *Clinical Infectious Diseases*, 33, 1004–1009.

Thomas, D. G., and Gart, J. J. (1977), "A Table of Exact Confidence Limits for Differences and Ratios of Two Proportions and Their Odds Ratios," *Journal of the American Statistical Association*, 72, 73–76.

Trollfors, B., Taranger, J., Lagergard, T., Sundh, V., Bryla, D. A., Schneerson, R., and Robbins, J. B. (1998), "Immunization of Children With Pertussis Toxoid Decreases Spread of Pertussis Within the Family," *The Pediatric Infectious Disease Journal*, 17, 196–199.

World Health Organization (1991), Meeting on Case Definition of Pertussis, January 10–11 1991, MIM/EPI/PERT/91.1.

World Health Organization (2001), Pertussis Surveillance: A Global Meeting, Geneva, October 16–18 2000, WHO/V and B/0.1.19.

Wolfinger, R. D. (1999), "Fitting Nonlinear Mixed Models With the New NLMIXED Procedure," in *Proceedings of the 24th Annual SAS Users Group International Conference*, Cary, NC: SAS Institute, Inc., p. 287.