Estimating vaccine efficacy in stochastic SIR epidemic models with non homogeneous mixing.

Carlos M. Hernández-Suárez
Carlos Castillo-Chavez

Summary

A stochastic SIR epidemic model with a "leaky" vaccine is developed, and it is shown that if the population can be divided in groups such that there is homogeneous mixing within groups, then, regardless of the mixing preferences among them, the amount of exposure to infection is the same for every individual within a given group, although individuals of different groups may have different amount of exposure. As a consequence, the usual method to estimate vaccine efficacy (VE) can be applied to each group and a weighted estimate of VE can be constructed. It is not required that all groups have the same size or the same contact and removal rates, and the fraction of vaccinated may be different for every group. The method does not require final attack data, nor information on the mixing proportions between groups. Stochastic simulations prove the feasibility of the method. The results can be considered the stochastic analog of those of Haber *et. al.* (1995a).

Keywords: *Vaccine efficacy, Vaccination, General epidemic model.*

1 Introduction

In the fight against disease there is no doubt that prevention has always been the best alternative, and for many diseases (measles, polio, influenza), vaccination is the only practical alternative to reduce an individual's susceptibility to infection. Evaluation of the protective effects of a vaccine is generally done by standardizing the amount of exposure of the vaccinated and the unvaccinated, and measuring some form of the effects of the disease in both groups. Several indexes have been suggested, not all of them measuring the same quantity. The range of conditions required to apply these methods is very broad, for instance, some methods can only be applied at the end of the outbreak, others require a specific underlying transmission model, or the knowledge of the distribution of the latent and infective period, etc., or at least some of its moments.

In the transmission of the disease, it is assumed that an individual makes contacts at the points of a Poisson process with rate λ. These models assume that infection of a susceptible occurs with probability one when this has contact with an infective, which can be modified to include a probability of transmission β that could depend for instance on a preventive measure such as vaccination. The parameter for the rate of infection becomes then $\lambda\,\beta$.

A vaccine could have a heterogeneous effect in the population, and thus β could have a distribution with mean β_0 for the unvaccinated and β_1 for the vaccinated. The different types of effects of a vaccine give raise to different models, for instance, when all vaccinated people respond equally to the vaccine then the vaccinated individuals have susceptibility β_1. If $\beta_1 > 0$ then the vaccine is called "leaky", or model 1 (Halloran *et al.*, 1992a). It would be possible also that the vaccine offers complete protection to all vaccinated, then $\beta_1 = 0$, in addition, some of the vaccinated individuals could have susceptibility β_1 while others a susceptibility β_2, $\beta_1 \neq \beta_2$. If $\beta_1 = 0$ and $\beta_2 = \beta_0$ then the vaccine is completely effective in a fraction of the vaccinated.

Halloran *et al.* (1992a) called this the "all-nothing" (1/0) vaccine model or model 2. When $\beta_0 > \beta_1 > 0$ and $\beta_2 = \beta_0$, then the vaccine has no protection in a group and a "leaky" effect in the other, this model is the "leaky/nothing" or model 3. When $\beta_0 > \beta_2 > 0$ and $\beta_1 = 0$, then the vaccine has complete protection in a group and a "leaky" effect in the other, this model is the "all/leaky" or model 4. Finally, when $0 < \beta_1 \leq \beta_2 \leq \beta_0$ the model is called "general" or model 5. For more discussion of these models see Halloran *et al.* (1991,1992a), Haber *et al.* (1991a), Farrington (1992), and Longini *et. al.* (1993b).

Vaccines may also have a waning effect, which makes the estimate to be dependent on time. For estimation of VE in the presence of waning see Durham *et. al.* (1997). Also, the vaccine may not only reduce the susceptibility of a vaccinated individual to infection, but also could reduce the infectiousness of an infective, as it is the case of oral polio vaccine and *Haemophilus influenza* type b vaccine, affecting the transmission of the infectious agent (Longini *et. al.*,1996). An example of a vaccine that affects only transmission is the malaria transmission blocking vaccine (Halloran *et. al.* 1992b). For estimation of VE for susceptibility and infectiousness see Longini *et. al.* (1996).

Several quantities provide information on the effects of a vaccine. Assume that a fraction f of the population is vaccinated and that a proportion α_i of the vaccinated have susceptibility β_i. One important parameter to estimate is $\overline{\beta} = \sum_i \alpha_i \beta_i$, the average susceptibility of vaccinated individuals, as well as the average vaccine efficacy, $\overline{\beta}/\beta_0$, also called summary vaccine efficacy. A different measure of the effect of vaccination is given by the *population effectiveness of vaccination*, which is the fraction of the disease cases prevented by a vaccination campaign, that is the ratio of the attack rates in the vaccinated population and the expected attack rate in the unvaccinated population. Since once a fraction of the population has been vaccinated the unvaccinated individuals are provided with herd immunity, the information provided by the unvaccinated individuals on the effects of non-vaccination is biased. Significative advances in

this problem under non homogeneous mixing have been done by Haber *et. al.* (1995b, 1997). See also Halloran *et. al.* (1996).

Here we deal with model 1 and 2 vaccines, and use the measure of vaccine efficacy defined by Haber *et. al.* (1991a).

$$VE = 1 - \beta_1/\beta_0$$

which standardizes for exposure among vaccinated and unvaccinated individuals.

Halloran *et. al.* (1992a) gives an account of some of the problems that arise in the estimation of vaccine efficacy. Among the assumptions usually required in evaluation of vaccine efficacy, homogeneous mixing of the population is perhaps the most difficult to fulfill. Although it could be realistic to assume homogeneous mixing at some level, like households, neighborhoods or schools, it is not likely that this assumption will hold between groups. This imposes a problem when estimating vaccine efficacy since some individuals will be subject to higher exposure than others and it may be possible that vaccinated individuals have had higher or lower exposure to the infectious agent, and thus the estimate of VE is biased. In an attempt to uniformize for exposure, the VE can be estimated using household secondary attack rates (SAR's) or with methods that require knowledge of the mixing proportions. In this paper we deal with this problem and show that good estimates can be constructed as long as groups of individuals that mix homogeneously can be identified.

The outline of the paper is as follows: in section 2 current methods to estimate vaccine efficacy are reviewed. In section 3 a different construction of an SIR epidemic model due to Sellke (1983) is presented. Based in this construction, we derive a relationship between the number of susceptibles and the current severity of the disease or accumulated exposure, and show that this holds for a more general class of epidemic models and conditions. Using this relationship, we derive the classical estimates of vaccine efficacy in section 2. In section 4, current methods to estimate vaccine efficacy for heterogeneous mixing are reviewed, whereas in

section 5 the results of section 3 are extended to this situation. Section 6 presents results of stochastic simulations.

2 Current results for estimating vaccine efficacy with homogeneous mixing.

Longini *et al.* (1993b) described five desirable properties in a vaccine efficacy (VE) estimate:

(*i*) have a clear and meaningful biological interpretation,

(*ii*) measure only the direct biological effect of immunization on vaccinated persons,

(*iii*) be theoretically invariant across different populations and study designs,

(*iv*) provide comparable exposure to infection for the unvaccinated and vaccinated, and

(*v*) standardize the exposure to infection in the unvaccinated and vaccinated.

There are several levels of information that one can have on an epidemics, mainly event time data or attack data. Rhodes *et al.* (1996) classify the information in more levels depending on the detail. According to that classification, we deal here with level IV information in which we only know whether an infection occurred or not to each individual in the population in some time period.

For level IV data, as described by Longini *et. al.* (1993b) there are basically three measures of vaccine efficacy: those based in the attack rates, those based in transmission rates VE(β) and those based in household secondary attack rates VE(SAR) or household transmission rates VE(β^*). Whether the outbreak is over or not defines if the level IV data corresponds to final

attack data or not. We use a slight modification of the notation in Longini *et al.* (1993b), to consider the possibility of estimating the VE before the outbreak is over:

n_0 number of unvaccinated in the population.

$X_0(t)$ number of unvaccinated who get the disease by time t.

n_1 number of vaccinated in the population.

$X_1(t)$ number of vaccinated who get the disease by time t.

$AR_0(t) = X_0(t)/n_0$ attack rate in the unvaccinated by time t.

$AR_1(t) = X_1(t)/n_1$ attack rate in the vaccinated by time t.

Hereafter, whenever the time index t is not specified, indicates that the value corresponds to the value at the end of the outbreak.

The first measure of vaccine efficacy was suggested by Greenwood and Yule (1915) with the equation

$$\text{VE} = \frac{AR_0 - AR_1}{AR_0} = 1 - \frac{AR_1}{AR_0}, \tag{3.1}$$

which is the classical measure based in the attack rates. The main drawback of (3.1) is that it measures the overall effect of a vaccine in a vaccinated population under a particular set of circumstances, which will make it vary with the degree of transmission in the community, for instance, it is known that the vaccine has indirect effects on the susceptibility of the population by reducing the number of infectives (herd immunity) and thus in this situation the efficacy of a

vaccine will be overestimated. In other words, attribute (v) will not be satisfied. Longini *et. al.* (1993b) showed that (3.1) increasingly underestimates β_1/β_0 with the level of transmission.

Vaccine efficacy based in transmission rates are based either in solving deterministic system of equations (Haber *et. al.*, 1991a, Longini *et. al.*, 1995) or in its stochastic analog (Becker, 1982). The measure based in the deterministic model is as follows:

Let $S_v(t)$, $I_v(t)$ and $R_v(t)$ be respectively the number of susceptible, infective and removed individuals with vaccination status v at time t, with $v = 0, 1, 2, ..., m$. Let α_v be the proportion of individuals in stratum v. The system of differential equations describing the deterministic model is:

$$S'_v(t) = -\lambda \beta_v(t)\ I_v(t)\ S_v(t)/n, \tag{3.2}$$

$$I'_v(t) = \lambda \beta_v(t)\, I_v(t)\ S_v(t)/n - \gamma\, I_v(t), \tag{3.3}$$

$$R'_v(t) = \gamma\, I_v(t), \tag{3.4}$$

with initial conditions $S_v(t) = n_v$, $R_v(t) = 0$, $I_v(t) > 0$. The parameter $\theta_v = \beta_v/\beta_0$ is the relative susceptibility of individuals vaccinated in stratum v against the unvaccinated, and the summary relative susceptibility is $\overline{\theta} = \sum_v \alpha_v\, \theta_v$.

The attack rate by time t is

$$AR_v(t) = 1 - S_v(t)/n_v. \tag{3.5}$$

Substituting (3.4) in (3.2) and evaluating at (3.5) gives

$$\beta_v = -\frac{\gamma}{n\lambda R(t)} log(1 - AR_v(t)), \tag{3.6}$$

for a leaky vaccine $v = 1$ and thus (3.6) is evaluated at the defined vaccine efficacy $1 - \beta/\beta_0 = 1 - \theta$, yielding

$$V\widehat{E}(\theta) = 1 - \frac{log\left(1 - A\widehat{R}_1(t)\right)}{log\left(1 - A\widehat{R}_2(t)\right)}, \tag{3.7}$$

where $A\widehat{R}_1(t)$ and $A\widehat{R}_2(t)$ are estimates of the attack rates of the epidemics at time t, (the current epidemic is a realization). Since they are binomial proportions,

$$A\widehat{R}_v(t) \xrightarrow{D} N[\, AR_v(t), AR_v(t)(1 - AR_v(t))/n_v \,]. \tag{3.8}$$

Using the delta-method,

$$V\widehat{E}(\theta) \xrightarrow{D} N\left[VE(\theta), \left(AR_1(t)/n_1 + \theta^2 AR_0(t)/n_0\right) [\, log(1 - AR_0(t)) \,]^{-2}\right] \tag{3.9}$$

as $n_v \to \infty$ and $I(t)$ large. Therefore, an estimate of the variance of VE(θ) is:

$$\left(A\widehat{R}_1(t)/n_1 + \theta^2 A\widehat{R}_0(t)/n_0\right) [\, log\left(1 - A\widehat{R}_0(t)\right) \,]^{-2}. \tag{3.10}$$

Test of hypothesis and confidence intervals can be derived from (3.9) and (3.10).

Becker (1982) and Longini *et. al.* (1993a) derived similar results to (3.7)-(3.10) using a Martingale approach.

3 An alternative construction of an SIR epidemic with homogeneous mixing.

Sellke (1983) derived an elegant yet simple construction of the stochastic SIR epidemic model. His goal was to provide a simple proof of Daniels' result (1967) regarding the distribution of the final number of removed, the "size" of the epidemics. Daniels stated that under the assumption that the initial number of infectives was small, then '...when the threshold is large but the population size is much larger, the distribution of the number remaining uninfected in a large epidemic has approximately the Poisson form with deterministic mean $Ne^{-N/\rho}$. ' Here, $\rho = N\mu\lambda^{-1}$.

Sellke's derivation relaxed the requirement of a small number of initially infectives, but still needed the assumption of exponentially distributed infectious period. Here, Ball (1986) version of Sellke's construction is preferred, since it does not require the assumption of exponential distribution of the infectious state.

Suppose that a initially infectives are introduced in a population of N susceptibles. Let $L_1, L_2, \cdots L_N$ be independent an identically distributed exponential random variables with parameter λ/N. Let $T_{-(a-1)}, T_{-(a-2)}, \cdots, T_N$ be a sequence of independent and identically distributed random variables, each distributed according to T_I, the duration of the infectious period. Let $L_{(1)}, L_{(2)}, \cdots L_{(N)}$ be the order statistics of $L_1, L_2, \cdots L_N$. Then the epidemic can be constructed as follows: For $i = -(a-1), -(a-2), \cdots, 0$ the initially infective individual i remains infective for a time T_i, after which it is removed. The i-th susceptible individual accumulates '*exposure to infection*' at a rate equal to the number of infectives present. When the total exposure of infection of individual j reaches L_j then individual j becomes infected. The j-

th individual so infected remains infective for time T_j and is then removed. The epidemic terminates with exactly n initial susceptibles infected when

$$\sum_{i=-(a-1)}^{n} T_i < L_{(n+1)},$$

and the total size of the epidemic R is given by

$$R = \min\left\{n : \sum_{i=-(a-1)}^{n} T_i < L_{(n+1)},\ n \geq 0\right\},$$

where the total area under the trajectory of infectives, W, is given by $W = \sum_{i=-(a-1)}^{R} T_i$.

The explanation of the construction above can be facilitated with the aid of Fig. 1:

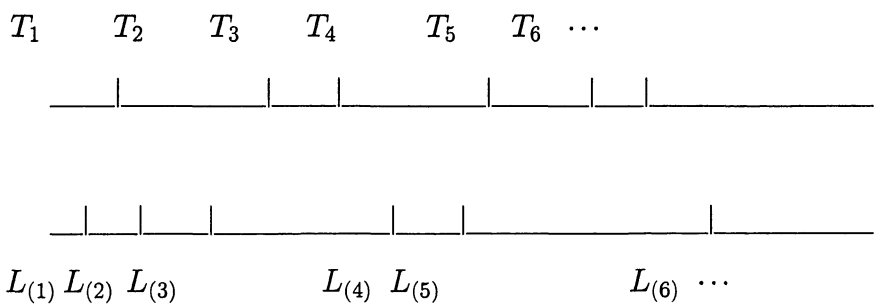

Fig. 3.1 Construction of an SIR epidemic model. The intervals in the top line correspond to the duration of the infectious state of individuals and the marks in the line of the bottom the amount of severity required to infection.

In Fig. 1, the intervals on the line on the top correspond to the durations of the infectious period of the individuals in the order they became infected, and the line in the bottom represent the exposure required to achieve infection.

Thus, the first individual requires a *'exposure to infection'* equal to $L_{(1)}$. Since (in Figure 1) this value is smaller than T_1, the amount of severity required for the first susceptible individual to become infective can be provided by the initial infective. Now, the second infection can be achieved if a susceptible individual is exposed to $L_{(2)}$ units of severity, and since this value is smaller than $T_1 + T_2$, there is enough severity to get a second infection, and so on. Then the epidemics terminates with R cases when the $R+1$ th infection requires more than $T_1 + T_2 + \cdots T_R$ units of exposure. The total exposure produced in this epidemics corresponds to the severity. In Fig. 1 the size of the epidemics (excluding the initial infective) is 5, since $L_{(6)} > \sum_{i=1}^{6} T_6$. This method of construction proves to be very fast in the simulation of SIR stochastic epidemic processes.

Some remarks

The epidemic can be constructed similarly by assuming that there is only one infected along the duration of the epidemic, who "marks" individuals (instead of infecting them) and that every time that this occurs, the infective extends its infection period for a random amount T_j. Thus, all of the remaining individuals have had the same *amount of exposure to the infected.* After some fixed time t, the amount of unmarked individuals (susceptibles) can be used to estimate the current amount of severity of the epidemics. Define $W(t)$ the amount of severity by time t, and define $W(\infty) \equiv W_\infty$, clearly, if the epidemics is over at some time t then $W(t) = W_\infty$. Observe that, by construction, the number of observations greater than a fixed value of severity w is

$$\text{Binomial}\left(N, e^{-\lambda w/N}\right),$$

therefore

$$E\{s_i(t)/n_i\} = 1 - Exp[-\lambda w/N] \tag{3.11}$$

which can be used to make inferences on $\lambda W(t)/N$.

The above model assumes that contacts between an infective and a susceptible results in immediate infection of the susceptible. We can include here a fixed probability that such a contact will result in infection of the susceptible. Consider a "leaky" vaccine that divides the population on n_1 vaccinated (status 1) and n_0 unvaccinated (status 0). Since there is homogeneous mixing, all individuals have had the same amount of "exposure to the infective", namely $W(t)$. Hence $s_i(t)$, the number of susceptible individuals with vaccination status i by time t follows a Binomial distribution with parameters n_i and $e^{-\lambda\beta_i W(t)/N}$, $i = 0,1$.

Svensson (1991,1994) found that for large populations the following balance equation holds

$$I(t)/N = 1 - Exp[-\gamma W(t)], \tag{3.12}$$

where γ is the susceptibility of infection. We can see that $S(t)$ being a random quantity, the equality holds for the expectation in the left side.

Observe that by the construction that led to (3.11), upon conditioning in a given severity, the fate of a particular individual in the population is independent of that of the other, which leads to conclude the conditional independence of $s_i(t)$ given $W(t)$. Since $\mathrm{E}\{s_i(t)/n_i|$

$W(t)\} = e^{-\lambda\beta_i W(t)/N}$, the natural choice to estimate the vaccine efficacy $\mathrm{VE} = 1 - \theta = 1 - \beta_1/\beta_0$ is

$$1 - \frac{log(s_1(t)/n_1)}{log(s_0(t)/n_0)}, \tag{3.13}$$

which is precisely (3.7). Observe that (3.13) is the maximum likelihood estimate of $1 - \theta$. By the conditional independence of the random variables involved

$$\mathrm{E}\{\widehat{\theta}\} = \mathrm{E}\left\{\frac{log(s_1(t)/n_1)}{log(s_0(t)/n_0)}\right\} \leq \frac{\mathrm{E}\{log(s_1(t)/n_1)\}}{\mathrm{E}\{log(s_0(t)/n_0)\}} = \frac{\lambda\beta_1 W(t)/N}{\lambda\beta_0 W(t)/N} = \frac{\beta_1}{\beta_0}$$

which is an alternative proof of the result of Haber *et. al.* (1991a) in the sense that $1 - \widehat{\theta}$ overestimates the vaccine efficacy.

4 Current methods to estimate efficacy under non homogeneous mixing.

The methods presented in this section, based in the secondary attack rates (SAR) can be applied also under homogeneous mixing. In the absence of homogeneous mixing, current estimates of vaccine efficacy are based on household secondary attack rates, the number of infections caused by an infective in a household. The idea behind this approach is as follows: a susceptible individual in the population could be infected from other member of his household or from other member of the community. If one could estimate the household probability of infection for both vaccinated and unvaccinated individuals then we could construct an estimate of efficacy of a vaccine using the ratio of both probabilities.

Denote by γ_{ki} the probability that during a short time unit a susceptible person from stratum k and vaccination status i becomes infected from a *single* infected household member. If there are two kinds of individuals, unvaccinated (group 0) and vaccinated (group 1), the vaccine efficacy in stratum k is defined as

$$\phi_k = 1 - \gamma_{k1}/\gamma_{k0}.$$

Define SAR_{ki} the probability of a susceptible in stratum k and vaccination status i becoming infected from a single infected person in the household during the duration of his infectious period. Then

$$\gamma_{ki} = 1 - (1 - \text{SAR}_{ki})^{1/\tau}$$

where τ is the average duration of the infectious period. The estimate of vaccine efficacy is then

$$\widehat{\phi}_k = 1 - \widehat{\gamma}_{k1}/\widehat{\gamma}_{k0} = 1 - \frac{1 - (1 - \text{SAR}_{k1})^{1/\tau}}{1 - (1 - \text{SAR}_{k0})^{1/\tau}}. \tag{3.15}$$

In Longini *et al.* (1993b) a review of the methods to estimate the SAR is made. These are: the conventional method, the model-based method of Longini *et al.* and the model-based method of Rampey *et al.* We briefly review these methods following the notation of Longini *et al.* (1993b). Let

β_0^* = transmission rate within the household to an unvaccinated susceptible, where $\beta_0^*(t)\Delta t + o(\Delta t)$ is the probability that an unvaccinated susceptible is infected due to contact with an infected person in Δt units of time within the household.

β_1^* = transmission rate within the household to an unvaccinated susceptible.

SAR_0 = household secondary attack rate to an unvaccinated susceptible.

SAR_1 = household secondary attack rate to a vaccinated susceptible.

Method I. The conventional method.

In this method, the researcher makes decisions about who is secondary to who among the infections in a household. Clearly, a knowledge on the duration of the latent, incubation and infectious period of the disease is required. Fixed intervals for these states are established and secondary cases are identified according to if they occurred in a given state of the index case. As mentioned by Longini *et al.* (1993b), the method presents some drawbacks, for instance misclassification errors and that it ignores co-primary cases.

Method II. The model-based method of Longini et al.

This method was developed by Longini and Koopman (1982, 1988) and requires final value data. Define CPI as the community probability of infection (infection from a member outside the household). They showed that the probability that exactly j persons become infected in a household with s initial susceptibles is

$$\pi_{js} = \binom{s}{j} \pi_{jj} (1 - \text{CPI})^{s-j} (1 - \text{SAR})^{j(s-j)}$$

where $\pi_{ss} = 1 - \sum_{j=0}^{s-1} \pi_{js}$, an expression that allows for maximum likelihood estimation of CPI and SAR without having to identify primary and secondary cases.

Method III. The model-based method of Rampey et al.

Rampey *et al.* (1992) developed a model that uses all information on infections and does not require that the epidemics is over. The method requires the specification of the probability distributions for the latent, incubation and infectious period.

In general, when there is illness incidence data available, one can estimate β_i^* given the relationship between this and the SAR_i:

$$\text{SAR}_i \approx 1 - (1 - \beta_i^*)^{\tau_1} \tag{3.16}$$

where τ_1 is the expected value of the length of the infectious period. Substituting this into (3.15) yields

$$VE(\text{SAR}) = 1 - \frac{1-(1-\beta_1^*)^{\tau_1}}{1-(1-\beta_0^*)^{\tau_1}}$$

Longini *et al.* (1993b) stated that if $\tau_1 \neq 1$, VE(SAR) $\neq \beta_1/\beta_0$ and hence both measures of efficacy are different. As mentioned by Longini *et al.* VE(SAR) and β_1/β_0 are both measures of efficacy over different exposure times. Clearly β_1/β_0 compares the probabilities of infection per contact between a susceptible and an infective.

Of these three methods, Method II seems more plausible due to the possibility of misclassification of method I and the need of assuming distributional properties of method III. Still, method II requires knowledge on the expected duration of the infectious period, and can only be applied to final attack data.

Haber *et al* (1995a) derived the following result for a deterministic SIR epidemic model for a population that is divided in groups with non-random mixing between groups and random mixing within them:

$$A_i = 1 - Exp\left[-\tau\lambda_i \sum_j \rho_{ij}\beta_{ij} A_j \right] \tag{3.17}$$

where A_i is the attack rate in group i, τ is the length of the infectious period (which is assumed to be the same for all groups), λ_i the contact rate of group i, ρ_{ij} is the proportion of contacts that a person of group i makes with a person of group j, and β_{ij} is the proportion of contacts between a susceptible from group i and an infective from group j that result in a new infection. For this deterministic model, Haber *et. al.* (1995a) showed that if vaccination does not affect the contact rates, then an estimate of the vaccine efficacy in group k is

$$VE_k = 1 - \frac{log(s_{k1}(t)/n_{k1})}{log(s_{k0}(t)/n_{k0})} \tag{3.18}$$

regardless of the mixing proportions between groups. Observe that if groups are of size one and there is homogeneous mixing at rate λ, then $\lambda_i \rho_{ij} = \lambda/N$. If $\beta_{ij} = \beta \, \forall \, i,j$, then we have

$$A_i = 1 - Exp\Big[- \tau\lambda\beta \sum_j A_j/N \Big]$$

since $A_i = 0, 1$, the term in the exponent is the severity of the disease at the end of the outbreak times a constant. Observe that if τ is random then this is (3.11). With the results of section 3.2, the stochastic analog of (3.17) is straightforward.

5 Estimation of the VE for a leaky vaccine under non homogeneous mixing.

The main conclusion of section 3 was that we can construct the maximum likelihood estimate of the ratio $\theta = \beta_1/\beta_0$ as long as both types of individuals, vaccinated and unvaccinated have had the same amount of exposure to infection, which is guaranteed under homogeneous mixing. When the population is divided in strata such as neighborhoods, schools, households etc., then we still can estimate β_1/β_0 for any subgroup of the population as long as the individuals in the group have had the same amount of exposure. We need to add a subscript to

the estimate $\widehat{\theta}$ to refer to stratum k, $k = 1, 2, \cdots, m$. Let n_{ki} be the number of individuals in group k with vaccination status i, $i = 0, 1$, and $n_k = \sum_i n_{ki}$ the total size of group k, then the MLE of $1 - \beta_1/\beta_0$ for stratum k when both types of individuals have had the same amount of exposure to infection is

$$1 - \frac{log(s_{k1}(t)/n_{k1})}{log(s_{k0}(t)/n_{k0})}. \tag{3.19}$$

Hereafter we use the term *group* for disjoint subsets of the population of individuals such that

a) the individuals of group k have a contact rate λ_k and mix homogeneously among them, and
b) the proportion of contacts that an individual of group i has with a member of group j is q_{ij}.

Let $v_k(t)$ be the accumulated number of infections in group k by time t, and let $x_{ki}(t)$ be the duration of the infectious period of individual i in stratum k up to time t. The severity of stratum k up to time t is then

$$W_k(t) = \sum_{i=1}^{v_k(t)} x_{ki}(t), \ k = 1, 2, \ldots, m.$$

The probability that an individual with vaccination status i survives infection up to a given time t is the probability that s/he survives infection from all infected in all groups. For the sake of simplicity, assume as in section 3 that contacts between infective and a susceptible will

result in immediate infection, thus the probability that an individual in stratum k will not meet with an infective in stratum j is $Exp[\ -\lambda\, q_{kj} W_j(t)/n_j]$.

Again, we use an analogy similar to that of section 3, where instead of infecting individuals, there is at most one single infective in every group. This infective "marks" individuals in his own group but can produce the first infection in other groups. Infections in his own group extends his infection period by a random amount x_k corresponding to group k, whereas infections produced in other groups add a random amount to the newly infected, (not necessarily with the same distribution). Using this construction, the relationship between the number of susceptibles in every group and the accumulated severity in the whole population is easier to derive.

First, consider group k with n_k initial susceptibles. Since the probability that a susceptible in this group meets an infective in any group (including his own) are independent events, the probability that an individual of this group is still susceptible at time t is the product of the probabilities that he did not meet the infectives in all groups by time t, which is given by

$$Exp\left[\ -\lambda\ q_{kj} \sum_j W_j(t)/n_j \right].$$

Incorporating now the vaccination effect on the i-th individual in group k, the probability becomes

$$Exp\left[\ -\lambda\ q_{kj}\beta_{ki} \sum_j W_j(t)/n_j \right].$$

Assume that the vaccine does not affect the contact rates of vaccinated individuals, then

$$\lambda\, q_{kj} \sum_{j} W_j(t)/n_j$$

is the same for all individuals in the k-th group, that is, once a group has an infective, except for the term β_{ki} all of the remaining $n_k - 1$ members of a group have had the same *amount of exposure to the initial infected* in his own group as well as to the initial infective in the other groups, if any. Therefore, the number of susceptibles in a group at an arbitrary but fixed time t provides information on the total amount of exposure at which individuals of that group have been subjected. The exposure rates of both vaccinated and unvaccinated in a given group k should differ only by the effect of vaccine β_{k1}, thus the distribution of the number of susceptibles with vaccination status i in stratum k, that have survived a severity w_j from stratum j by time t is

$$S_{ki}(t) = Bin\left(z_{ki},\, Exp\Big[-\lambda \beta_{ki}\, q_{kj} \sum_{j} w_j/n_j \Big] \right), \tag{3.20}$$

where $z_{ki} = z_{ki} - 1$ if the index case in this group was of vaccination status i and $z_{ki} = n_{ki}$ otherwise, that is, similarly to the SAR's, we do not consider the index case in every group.

The expected number of infected obtained from (3.20) is the stochastic analog of (3.17) since in that result τ, the time of infection is assumed constant for all individuals in all groups. It can be seen in (3.20) that it is the total severity what matters, which implies that (3.20) holds independently of the distribution of the duration of the infectious state, and it may even be different for every group or every individual. This may be useful if one wishes to consider the possibility of some groups (or individuals) being more efficient in the removal or isolation of infectives, which may be a realistic situation if removal or isolation is affected by socioeconomic status.

In general, whatever the contact structure of the population, if this can be divided in groups with the defined properties, then the individuals of any group have the same amount of exposure, although individuals of different groups may have different exposure. If there is a group related vaccine effect, that is, if β_i becomes β_{ki} where k is the stratum index, then (3.19) still gives the specific VE for every stratum. In Haber Longini *et. al.* (1991b) a situation is given in which there is a population of individuals mixing homogeneously but the population is divided in k strata with related vaccine efficacy θ_k for stratum k. Observe that these strata meet the conditions *a* and *b* that define a group, therefore θ_k can be estimated via (3.19).

If there is no vaccine related effect then we can construct an estimate of the vaccine efficacy across groups with

$$VE^* = 1 - \sum_{u=1}^{k} w_u \widehat{\theta}_u \tag{3.21}$$

with $\widehat{\theta}_u$ calculated as in (3.19) and w_u is the weight given to the estimate in group u. With random vaccination there is no reason why the number of vaccinated and unvaccinated susceptibles in every group should be equal, thus using the group sizes as weight is discarded. A natural choice for the w_u would be

$$w_u = \frac{\sigma_i^{-2}}{\sum_i \sigma_i^{-2}}$$

the reciprocal of the normalized variance in group u, as in Haber *et al.* (1991b), see also Casella (1990) p.338) although it must be remembered that the estimates are not unbiased . In practice we would substitute an estimate of this variance. The estimated variance of $\widehat{\theta}$ in (3.21) becomes

$$\widehat{Var}(\widehat{\theta}) = \left(\sum_i \sigma_i^{-2} \right)^{-1}$$

If there is a stratum associated susceptibility, for instance age, then there is a constant multiplying β_1/β_0 that becomes absorbed by the contact rate, therefore, (3.21) still provides an estimate of θ.

6 Simulations

In order to test the estimate the estimate (3.21), some stochastic simulations of SIR epidemics were performed. The population has 560 individuals in three groups, of sizes $n_1 = 300$, $n_2 = 60$ and $n_3 = 200$. The contact rates for each group are: $\lambda_1 = 1.5$, $\lambda_2 = 7.5$ and $\lambda_3 = 4.5$. The mixing proportions are given by the matrix Q:

$$Q = \begin{bmatrix} 0.5 & 0.3 & 0.2 \\ 0.3 & 0.6 & 0.1 \\ 0.1 & 0.05 & 0.85 \end{bmatrix}$$

where $\{q_{ij}\}$ is the proportion of contacts that a person of group i makes with a person of group j. The matrix Q satisfy the balance conditions $n_i\lambda_i\, q_{ij} = n_j\lambda_j\, q_{ji}$, $i, j = 1, 2, 3$. In the simulations β_0 is set to one and β_1 varies as well as the fraction of vaccinated in each group.

The values of β_1 were 0.05,0.1,0.2, and 0.3, whereas three different vaccination schemes were tested, these are {0.6,0.4,0.5}, {0.7,0.4,0.2} and {0.2,0.5,0.7} where $\{f_1, f_2, f_3\}$ is the fraction of vaccinated in groups 1,2 and 3 respectively. Vaccination is random in every group.

With the combination of parameters above, three sets of simulations where run. In the first set (cases 1-3), the removal rate is equal for all groups and the data is collected at the end of the epidemics. In the second set of simulations (cases 4-6) every group has different removal rate, and the data is also taken at the end of the epidemic. The third set of simulations (cases 7-9) is similar to the previous except that the data is taken when the severity reaches 100.

The duration of the infectious state was assumed to be exponential with mean μ_k^{-1} for group k, whereas an individual in group k makes contacts according to a Poisson process of intensity λ_k. Thus, the process is simulated according to a continuous time Markov chain. Let $S_i(t)$ and $I_i(t)$ be the number of susceptibles and infectives in group i at time t. The removal rate when the state of the system is $\{S_i(t),\ I_i(t)\}$ is $\mu_i \sum_i I_i(t)$ whereas the infection rate is

$$\sum_i I_i(t) \sum_j (\beta_0 S_{0j}(t)/n_j + \beta_1 S_{1j}(t)/n_j),$$

where $S_{vj}(t)$ is the number of susceptible individuals in group j with vaccination status v, $v = 0, 1$.

For each combination of parameters 300 simulations were run using Matlab. The estimated VE using (3.21) as well as the estimated VE using ignoring the structure of the population using (3.13) were calculated. The weighted estimate using (3.21) is referred as VE*, whereas the estimate constructed without considering the structure of the population as in (3.13) is refereed as the crude VE. For every simulation VE* and the crude VE where calculated and their averages and standard deviations are shown. The code is in the appendix (A4). Results are shown in Tables 1-3.

The estimateVE* proved to be accurate in all cases. Both estimates consistently underestimated the true value of VE, and in general the bias decreased with the vaccine efficiency. The bias of the crude VE estimate was generally larger than that of VE*, specially in low vaccine efficacy situations. The crude VE estimate proved to be very good when the fraction of vaccinated was about the same in the three groups (cases 1 and 4), but it proved to perform very poorly in the other four cases, where the variation in the fraction of vaccinated was larger. In general, VE had smaller variance than VE*.

Heterogeneous removal rates did not affect both estimates significatively. The differences between the calculated VE* for the case of equal removal rates and that with heterogeneous removal rates had an average of 0.039, and it was within 3 decimal places in 10 of the 12 cases. For the crude VE it was within 2 decimal places. VE* should be preferred to VE, since there are no more assumptions to use the former than the required to use VE, except that groups are identifiable and thus estimates for every group can be obtained.

Estimation of VE before the end of epidemics (at a severity of 100) did not affect VE* of VE significatively. By comparing cases 5 and 8 it can be seen that for a true VE of 0.7, evaluating at W=100 corresponded to estimation of VE at about half of the total cases with respect to evaluation at the end of the epidemic.

7 Discussion

The method presented here tells that the current measures of VE based in transmission rates can be applied under non-homogeneous mixing as long as the population can be divided in groups that mix homogeneously among them regardless of the contact structure between groups.

TABLE 1 Results of stochastic simulations. Evaluation at the end of the outbreak.
mu1 = mu2 = mu3 = 1

Case 1)
f 1 = 0.6 f 2 = 0.4 f 3 = 0.5

		ATTACK RATES (*)							
		VE = 0.95		VE = 0.90		VE = 0.80		VE = 0.70	
group	ni	AR0	AR1	AR0	AR1	AR0	AR1	AR0	AR1
1	300	0.4429	0.024	0.495	0.062	0.569	0.149	0.624	0.248
2	60	0.952	0.128	0.973	0.317	0.993	0.625	0.996	0.799
3	200	0.8515	0.085	0.906	0.21	0.947	0.455	0.967	0.642
Weighted VE (†)		0.949 ± 0.022		0.895 ± 0.037		0.797 ± 0.045		0.698 ± 0.059	
Crude VE (‡)		0.945 ± 0.013		0.884 ± 0.021		0.768 ± 0.031		0.665 ± 0.04	

Case 2)
f 1 = 0.7 f 2 = 0.4 f 3 = 0.2

		ATTACK RATES (*)							
		VE = 0.95		VE = 0.90		VE = 0.80		VE = 0.70	
group	ni	AR0	AR1	AR0	AR1	AR0	AR1	AR0	AR1
1	300	0.496	0.029	0.528	0.067	0.583	0.155	0.62	0.249
2	60	0.973	0.138	0.98	0.326	0.992	0.613	0.989	0.799
3	200	0.958	0.132	0.963	0.263	0.974	0.507	0.972	0.677
Weighted VE (†)		0.948 ± 0.02		0.898 ± 0.032		0.797 ± 0.047		0.693 ± 0.063	
Crude VE (‡)		0.961 ± 0.01		0.923 ± 0.015		0.848 ± 0.022		0.775 ± 0.029	

Case 3)
f 1 = 0.2 f 2 = 0.5 f 3 = 0.7

		ATTACK RATES (*)							
		VE = 0.95		VE = 0.90		VE = 0.80		VE = 0.70	
group	ni	AR0	AR1	AR0	AR1	AR0	AR1	AR0	AR1
1	300	0.451	0.014	0.504	0.052	0.59	0.15	0.641	0.258
2	60	0.935	0.13	0.963	0.312	0.968	0.609	0.981	0.808
3	200	0.688	0.054	0.775	0.144	0.887	0.386	0.937	0.6
Weighted VE (†)		0.946 ± 0.037		0.898 ± 0.046		0.798 ± 0.081		0.69 ± 0.09	
Crude VE (‡)		0.911 ± 0.038		0.822 ± 0.042		0.616 ± 0.074		0.403 ± 0.1	

* Average of 300 simulations
† using (3.21)
‡ ignoring structure of population

TABLE 2 Results of stochastic simulations. Evaluation at the end of the outbreak.

mu1 = 1.5 mu2 = 0.6 mu3 = 1.0

Case 4)

f 1 = 0.6 f 2 = 0.4 f 3 = 0.5

		ATTACK RATES (*)							
		VE = 0.95		VE = 0.90		VE = 0.80		VE = 0.70	
group	ni	AR0	AR1	AR0	AR1	AR0	AR1	AR0	AR1
1	300	0.556	0.036	0.603	0.081	0.667	0.192	0.692	0.293
2	60	0.993	0.225	0.994	0.471	0.989	0.777	0.996	0.9
3	200	0.885	0.098	0.918	0.217	0.95	0.469	0.966	0.653
Weighted VE (†)		.947 ± 0.020		.901 ± 0.026		.798 ± 0.052		.697 ± 0.056	
Crude VE (‡)		.940 ± 0.014		.883 ± 0.021		.765 ± 0.036		.663 ± 0.037	

Case 5)

f 1 = 0.7 f 2 = 0.4 f 3 = 0.2

		ATTACK RATES (*)							
		VE = 0.95		VE = 0.90		VE = 0.80		VE = 0.70	
group	ni	AR0	AR1	AR0	AR1	AR0	AR1	AR0	AR1
1	300	0.588	0.038	0.627	0.089	0.679	0.198	0.702	0.297
2	60	0.996	0.246	0.999	0.488	0.996	0.785	0.999	0.895
3	200	0.965	0.135	0.971	0.287	0.974	0.517	0.981	0.685
Weighted VE (†)		.948 ± 0.020		.898 ± 0.026		.799 ± 0.039		.699 ± 0.058	
Crude VE (‡)		.955 ± 0.010		.912 ± 0.015		.834 ± 0.023		.765 ± 0.029	

Case 6)

f 1 = 0.2 f 2 = 0.5 f 3 = 0.7

		ATTACK RATES (*)							
		VE = 0.95		VE = 0.90		VE = 0.80		VE = 0.70	
group	ni	AR0	AR1	AR0	AR1	AR0	AR1	AR0	AR1
1	300	0.536	0.021	0.587	0.07	0.677	0.199	0.716	0.292
2	60	0.98	0.208	0.983	0.443	0.996	0.785	0.997	0.907
3	200	0.732	0.059	0.812	0.155	0.923	0.402	0.963	0.631
Weighted VE (†)		.949 ± 0.025		.896 ± 0.038		.786 ± 0.061		.705 ± .064	
Crude VE (‡)		.913 ± 0.024		.822 ± 0.037		.625 ± 0.055		.428 ± 0.072	

* Average of 300 simulations

† using (3.21)

‡ ignoring structure of population

TABLE 3 Results of stochastic simulations Evaluation at a severity W=100

mu1 = 1.5 mu2 = 0.6 mu3 = 1.0

Case 7)

f 1 = 0.6 f 2 = 0.4 f 3 = 0.5

		ATTACK RATES (*)							
		VE = 0.95		VE = 0.90		VE = 0.80		VE = 0.70	
group	ni	AR0	AR1	AR0	AR1	AR0	AR1	AR0	AR1
1	300	0.516	0.03	0.507	0.063	0.477	0.116	0.431	0.153
2	60	0.936	0.176	0.894	0.327	0.799	0.432	0.728	0.47
3	200	0.832	0.081	0.807	0.156	0.735	0.253	0.656	0.313
Weighted VE (†)		.948 ± 0.019		.899 ± 0.031		.800 ± 0.046		.691 ± 0.064	
Crude VE (‡)		.941 ± 0.013		.885 ± 0.019		.782 ± 0.028		.681 ± 0.041	

Case 8)

f 1 = 0.7 f 2 = 0.4 f 3 = 0.2

		ATTACK RATES (*)							
		VE = 0.95		VE = 0.90		VE = 0.80		VE = 0.70	
group	ni	AR0	AR1	AR0	AR1	AR0	AR1	AR0	AR1
1	300	0.302	0.012	0.307	0.031	0.31	0.066	0.313	0.104
2	60	0.606	0.036	0.595	0.114	0.582	0.217	0.563	0.298
3	200	0.602	0.031	0.598	0.088	0.574	0.169	0.563	0.235
Weighted VE (†)		.949 ± 0.024		.895 ± 0.038		.797 ± 0.054		.691 ± 0.078	
Crude VE (‡)		.963 ± 0.011		.925 ± 0.017		.858 ± 0.024		.791 ± 0.032	

Case 9)

f 1 = 0.2 f 2 = 0.5 f 3 = 0.7

		ATTACK RATES (*)							
		VE = 0.95		VE = 0.90		VE = 0.80		VE = 0.70	
group	ni	AR0	AR1	AR0	AR1	AR0	AR1	AR0	AR1
1	300	0.526	0.02	0.549	0.067	0.558	0.138	0.515	0.187
2	60	0.976	0.207	0.945	0.38	0.892	0.552	0.816	0.582
3	200	0.722	0.057	0.766	0.135	0.768	0.269	0.72	0.348
Weighted VE (†)		.943 ± 0.089		.893 ± 0.043		.791 ± 0.055		.685 ± .072	
Crude VE (‡)		.909 ± 0.023		.818 ± 0.041		.639 ± 0.055		.448 ± 0.086	

* Average of 300 simulations

† using (3.21)

‡ ignoring structure of population

TABLE 4 Results of stochastic simulations. Evaluation at W = 70

mu1 = 1.5 mu2 = 0.6 mu3 = 1.0

Case 10)

f 1 = 0.6 f 2 = 0.4 f 3 = 0.5

		ATTACK RATES (*)							
		VE = 0.95		VE = 0.90		VE = 0.80		VE = 0.70	
group	ni	AR0	AR1	AR0	AR1	AR0	AR1	AR0	AR1
1	300	0.382	0.017	0.372	0.04	0.346	0.078	0.321	0.104
2	60	0.781	0.094	0.731	0.187	0.668	0.281	0.605	0.337
3	200	0.622	0.043	0.585	0.085	0.529	0.146	0.49	0.197
Weighted VE (†)		0.947 ± 0.039		0.897 ± 0.037		0.790 ± 0.060		0.694 ± 0.075	
Crude VE (‡)		0.943 ± 0.020		0.890 ± 0.023		0.786 ± 0.039		0.695 ± 0.041	

Case 11)

f 1 = 0.7 f 2 = 0.4 f 3 = 0.2

		ATTACK RATES (*)							
		VE = 0.95		VE = 0.90		VE = 0.80		VE = 0.70	
group	ni	AR0	AR1	AR0	AR1	AR0	AR1	AR0	AR1
1	300	0.202	0.006	0.208	0.017	0.212	0.041	0.217	0.064
2	60	0.455	0.003	0.451	0.05	0.447	0.013	0.439	0.189
3	200	0.428	0.005	0.414	0.039	0.403	0.09	0.392	0.13
Weighted VE (†)		0.949 ± 0.031		0.900 ± 0.040		0.793 ± 0.069		0.696 ± 0.083	
Crude VE (‡)		0.966 ± 0.012		0.931 ± 0.018		0.862 ± 0.028		0.798 ± 0.036	

Case 12)

f 1 = 0.2 f 2 = 0.5 f 3 = 0.7

		ATTACK RATES (*)							
		VE = 0.95		VE = 0.90		VE = 0.80		VE = 0.70	
group	ni	AR0	AR1	AR0	AR1	AR0	AR1	AR0	AR1
1	300	0.417	0.01	0.414	0.035	0.391	0.078	0.372	0.115
2	60	0.851	0.12	0.817	0.225	0.734	0.346	0.676	0.403
3	200	0.546	0.034	0.546	0.074	0.535	0.144	0.489	0.198
Weighted VE (†)		0.947 ± 0.029		0.897 ± 0.037		0.792 ± 0.065		0.687 ± 0.071	
Crude VE (‡)		0.912 ± 0.027		0.827 ± 0.040		0.660 ± 0.067		0.491 ± 0.089	

* Average of 300 simulations

† using (3.21)

‡ ignoring structure of population

It is expected that such a groups could be small (i.e. households) hence the usual asymptotic arguments used in 3.8-3.10. Nevertheless, for certain mixing patterns, individuals of different groups may have had approximately the same amount of exposure, and thus it is always possible to form larger groups. The notion behind a group is that they constitute the basic unit of homogenous mixing and thus, every individual in a group have the same "internal" amount of exposure to infection and whatever the distribution of the infectives outside the group and the mixing pattern between them, those individuals have the same amount of "external" exposure to infection. Larger groups can be formed by merging groups as well as every individual in the newly formed group has had the same amount of both "internal" and "external" exposure to infection. Let $\mathcal{A}$ be an arbitrary subset of groups. The probability that individual in $\mathcal{A}$ belonging to group i has survived infection by time t is $Exp[-U_i]$ where U_i is the amount of exposure to infection of individual in group i,

$$U_i = \sum_{k \notin \mathcal{A}} \lambda \, q_{ik} W_k(t)/n_k + \lambda \, q_{ij} W_j(t)/n_j + \lambda \, q_{ii} W_i(t)/n_i. \tag{3.22}$$

In some situations there are groups i,j such that $U_i = U_j$, for instance, assume that all groups have the same size, and consider two groups i and j in the arbitrary set $\mathcal{A}$. If $q_{ik} = q_{jk} \; \forall \, k \in \mathcal{A}^c$, that is, if individuals in groups i and j mix at equal probabilities with all groups outside $\mathcal{A}$, then since all groups have the same size $q_{ii} = q_{jj}$, therefore (3.22) holds for individuals in groups i and j as long as $W_i(t) = W_j(t)$, that is, if both groups have had the same amount of severity up to time t. If these conditions are met, then the exposure for both types of individuals in both groups is the same and therefore the attack rate data be merged.

Consider for instance the case of M neighborhoods, each one with h households each with n_k susceptibles. If the probability that an individual will contact an individual from other

house in the same neighborhood is the same for all houses within the neighborhood, then $q_{ij} = q_{ji}$. Thus we can group houses within neighborhoods if the assumption that the severity on each house is the same is valid. For final attack data, houses that had the same number of cases can be assumed to have had generated same amount of severity, and thus they can be merged in principle. Observe that no knowledge on the q_{ij}'s is required. More research on how to treat small groups is required.

One important assumption is that the contact rate of vaccinated individuals does not change, which is a somewhat difficult assumption in some cases. As previously stated, this affects the VE estimate by including indirect effects. Nevertheless, at least during the field trials of the vaccine it is sometimes possible to reduce this factor, for instance, with the use of placebos, thus, a potential vaccine effect on the contact rates affects equally vaccinated and unvaccinated individuals.

Finally, although here the VE is constructed with total attack rate data, adaptation for a sample of the population is straightforward.

8 References

[1] Ball, F.G. (1986) A unified approach to the distribution of total size and total area under the trajectory of infectives in epidemic models. *Adv. Appl. Prob.* 18:289-310.

[2] Becker, N.G. (1982) Estimation in models for the spread of infectious diseases. *Proc. XIth Int. Biometrics Conf.*,INRA,Versailles. pp.145-151,

[3] Becker, N.G. (1989) The Analysis of Infectious Disease Data, Chapman and hall, London.

[4] Casella, G. and Berger, R.L. (1990). Statistical Inference. Duxbury Press, California.

[5] Daniels,H.E. (1967) The distribution of the total size of an epidemic. *Proc. 5th Berkeley Symp. Math. Statist.Prob.* 4:281-293.

[6] Durham,L.K, Longini, I.M, Halloran, M.E, Clemens,J.D., Nizam A., and Rao,M. (1997) Estimation of vaccine efficacy in the prescence of wanning: application to cholera vaccines.Submitted to *Am. J. Epidemiol.*

[7] Farrington,C.P. (1992) The measurement and interpretation of age-specific vaccine efficacy. *Int. J. Epidemiol.* 21:1014-1020.

[8] Greenwood M. and Yule U.G. (1915) The statistics of anti-typhoid and anti-cholera inoculations, and the interpretation of suh statistics in general. *proc. R. Soc. Med.* 8(part 2):113-94

[9] Haber, M, Longini, I.M. and Halloran, M.E. (1991a) Measures of the effecs of vaccination in a randomly mixing population. *Int. J. Epidemiol.* 20:300-310.

[10] Haber, M, Longini, I.M. and Halloran, M.E. (1991b) Estimation of vaccine efficacy in outbreaks of acute infectious diseases. *Stat. Med.* 10: 1573-1584

[11] Haber, Halloran,M.E., and Longini, I.M. (1995a) Estimation of vaccine efficacy in non-randomly mixing populations. *Biometrical Journal.* 37:25-38

[12] Haber,M, Watelet,L and Halloran, E. (1995b) On individual and population effectiveness of vaccination. *Int. J. Epidemiol.* 24:1249-1260

[13] Haber, M. (1997) Estimation of the population effectiveness of vaccination. *Stat. Med.* 16:601-610

[14] Halloran, M.E., Haber,M, Longini,I.M. and Struchiner,C.J.(1991) Direct and indirect effects in vaccine efficacy and effectiveness. *Am. J. Epidemiol.* 133:323-331

[15] Halloran, M.E., Haber,M and Longini,I.M.(1992a) Interpretation and estimation of vaccine efficacy under heterogeneity. *Am. J. Epidemiol.* 136:328-343

[16] Halloran,M.E. and Struchiner,C.J. (1992b) Modeling transmission dynamics of stage-specificmalaria vaccines. *Parasitology Today.* 8:77-85.

[17] Halloran, M.E, Struchiner,C.J and Longini, I.M. (1996) Study design for different efficacy and effectiveness aspects of vaccines. Submitted to *Am. J. Epidemiol.*

[18] Longini, I.M., and Koopman,J.S. (1982) Household and community transmission parameters from final distributions of infections in households. *Biometrics* 38:115-126

[19] Longini, I.M., Koopman,J.S., Haber and Cotsonis, G.A. (1988) Statistical Inference for infectious diseases. *Am. J. Epidemiol.* 128:845-859

[20] Longini, I.M, Halloran, M.E.and Haber, M, (1993a) Estimation of vaccine efficacy from epidemics of acute infectious agents under vaccine-related heterogeneity. *Math. Biosc.* 117:271-281

[21] Longini, I.M, Halloran, M.E., Haber, M, and Chen, R.T. (1993b) Measuring vaccine efficacy from epidemics of acute infectious agents. *Stat. Med.* 12: 249-263

[24] Longini, I.M, Halloran, M.E.and Haber, M, (1995) Some current trends in estimating vaccine efficacy. In: Epidemic models: their structure and relation to data. Mollison,D. (Ed.) Publications of the Newton Institute. Cambridge University Press. pp. 394-403

[22] Longini, I.M, Sagatelian,K.,Rida, W., and Halloran, M.E. (1996) Optimal vaccine trial design when estimating vaccine efficacy for susceptibility and infectiousness from multiple populations.Submitted to *Stat. Med.*

[23] Rampey, A.H., Longini,I.M., Haber,M. and Monto,A.S. (1992) A discrete time model for the statistical analysis of infectious disease incidence data. *Biometrics 4*8:117-128

[25] Rhodes,P.R, Halloran,M.E and Longini, I.M. (1996) Counting process models for infectious disease data: distinguishing exposure to infection from susceptibility. *J.R. Statist. Soc.* 58:751-762

[26] Sellke,T.(1983) On the asymptotic distribution of the size of a stochastic epidemic.*J. Appl. Prob.* 20:390-394.

[27] Svensson, A. (1991) Analyzing the effect of vaccines. *Math. Biosc.* 107:407-412

[28] Svensson, A. (1994) On the simultaneous distribution of size and costs of an epidemic in a closed multigroup population. *Math. Biosc.* 127:167-180