

Urn models and vaccine efficacy estimation

Carlos M. Hernández-Suárez^{1,*,*†} and Carlos Castillo-Chavez²

¹ *CGIC - Universidad de Colima, Apdo. Postal 25, Colima, 28045 México*

² *Biometrics Unit, Cornell University, Warren Hall 435, Ithaca, NY 14850, U.S.A.*

SUMMARY

We derive the distribution of the number of infections among unvaccinated and vaccinated individuals for model 1 (leaky) and model 2 (all/nothing) vaccines, assuming random mixing of a homogeneous population. For all/nothing vaccines, we show that the distribution of the number of infected vaccinated individuals conditioning on n observed infections follows a hypergeometric distribution, and the vaccine efficacy estimate (VE) can be derived from the usual estimate of the total population size in a capture–recapture sampling program. For leaky vaccines, we show that the number of vaccinated infected follows a distribution that was first derived by Wallenius. We found that the current point estimates of VE for each model perform very well, but the urn model construction presented here provides a strong framework for estimation and hypothesis testing on the parameters, and can be applied when the available data are a sample of the population. Since the method does not require an underlying transmission model, it can be applied to estimate the VE for non-contagious diseases. Copyright © 2000 John Wiley & Sons, Ltd.

1. INTRODUCTION

Vaccines are designed to protect a population from infection. They work through direct and indirect effects, but usually it is the direct effect that we want to estimate. The first measure of vaccine efficacy was suggested by Greenwood and Yule [1] as

$$VE = 1 - \frac{AR_1}{AR_0} \quad (1)$$

where AR_0 and AR_1 are the attack rates among the unvaccinated and vaccinated individuals, respectively. O'Neill [2] derived the following approximation to the variance of $\ln(1 - VE)$ in (1):

$$\frac{(1 - AR_0)}{N_0 AR_0} + \frac{(1 - AR_1)}{N_1 AR_1} \quad (2)$$

If a vaccine confers total immunity to a fraction $1 - \beta$ of the vaccinated individuals and leaves susceptible the remaining vaccinated then it is called an ‘all/nothing’ or model 2 vaccine. If every

* Correspondence to: Carlos M. Hernández-Suárez, CGIC - Universidad de Colima, Apdo. Postal 25, Colima, 28045 México.

† E-mail: cmh2@cgic.ucol.mx

Contract/grant sponsor: CONACYT; contract/grant number: 438100-5-I26932-M

vaccinated individual reduces her/his chances of infection from a natural susceptibility β_0 to β_1 every time she/he becomes in contact with the infectious agent, then the vaccine is called 'leaky' or model 1 vaccine. Combinations of models 1 and 2 vaccine effects yield other models like 'leaky/nothing' or model 3 and 'all/leaky' or model 4. Here we are concerned only with models 1 and 2 vaccines. For a deeper discussion on the effects of vaccines see Halloran *et al.* [3,4], Haber *et al.* [5], Farrington [6] and Longini *et al.* [7]. For leaky vaccines we use the measure of VE defined by Haber *et al.* [5]:

$$VE = 1 - \frac{\beta_1}{\beta_0}$$

where β_1 is the susceptibility of a vaccinated individual and β_0 is the natural susceptibility of unvaccinated individuals conditional on exposure to infection. For all/nothing vaccines we use $VE = 1 - \beta$, that is, the proportion of fully protected individuals and we assume that unvaccinated individuals are susceptible to the disease.

Smith *et al.* [8] showed that if the vaccine decreases the probability of a disease given exposure (model 1) then the efficacy as in (1) will depend on time after beginning of exposure to infection. Thus, for leaky vaccines Haber *et al.* [5] suggested the estimator

$$VE = 1 - \frac{\ln(1 - AR_1)}{\ln(1 - AR_0)} \quad (3)$$

which has approximate variance [9]

$$\frac{AR_1/(N_1 - x) + (1 - VE)^2 (AR_0/(N_0 - n + x))}{(\ln(1 - AR_0))^2} \quad (4)$$

The estimators (1) and (3) can be obtained by writing a deterministic or a stochastic model. In both cases, a system of equations that explains how the epidemic is transmitted is solved for the parameters of interest. This method provides a way to derive point estimates of the VE although their statistical properties have not been derived. Also, since it is not clear what is the distribution of the number of infected among the vaccinated and unvaccinated, the method lacks a rigorous framework for interval estimation. In addition, stochastic models require some assumptions, for instance, contacts of individuals occur at the points of a Poisson process with parameter λ . It is natural to expect that both the contact rate and average infectious time will depend on time from the beginning of the outbreak, for instance, due to a learning process that has the effect of reducing the contact rate as well as the time of recovery (or removal) after infection. A model that includes these factors would be more realistic but obviously its analysis would be more involved and, due to the obvious need of extra assumptions, it is not clear if they will result in an increase in our knowledge of the VE.

In the next sections we derive the statistical distribution of the number of infected among vaccinated or unvaccinated for model 1 and 2 vaccines, and use these to construct estimators of VE for model 1 and 2 vaccines, keeping the amount of assumptions to a minimum. The type of action of the vaccine is assumed known.

2. INFERENCE FOR ALL/NOTHING (MODEL 2) VACCINES

Here $VE = 1 - \beta$, where β is the proportion of vaccinated that are susceptible to the disease and we assume that unvaccinated individuals are susceptible. Let the number of vaccinated and

non-vaccinated individuals be N_1 and N_0 , respectively, and x the number of infected among the vaccinated given an observed number of infections n (not necessarily at the end of the outbreak). Under random mixing in a homogeneous population: (i) the probability that the next infection will be a vaccinated individual depends only on the relative proportion of susceptible vaccinated to the total susceptible; (ii) an individual can be infected at most once. Thus, the number of infected among the vaccinated when a total of n infections is observed follows a hypergeometric distribution with parameters n , N_0 and $[N_1\beta]$ where $[z]$ represent the closest integer to z .

2.1. The case of a fixed number of protected individuals

In practice, if the vaccine is of type 2, then there is a probability $1 - \beta$ that a vaccinated individual will be fully protected, therefore Y , the total number of susceptible among the vaccinated, is a binomial random variable (N_1, β) . First we assume the number of unprotected vaccinated is a fixed quantity $[N_1\beta]$ and leave the random case for latter in this section.

As previously stated, at any stage of the epidemic if n infections are observed, the number of infected individuals among the vaccinated has probability mass function

$$P(X = x; n, N_0, N_1, \beta) = \frac{\binom{[N_1\beta]}{x} \binom{N_0}{n-x}}{\binom{N_0 + [N_1\beta]}{n}} \quad (5)$$

for $x = 0, 1, 2, \dots, \min(n, N_1)$. Since N_0 , N_1 and n are known, and x is observed, we can make inferences on β through inferences on the total population at risk $M = N_0 + [N_1\beta]$ as in a capture–recapture scheme, with N_0 being the number of ‘marked’ individuals and $n - x$ being the number of observed marked among the n recaptured.

Here we do not attempt to cover the extensive literature on the capture–recapture methods but just to outline some basic results; the maximum likelihood estimator (MLE) of the total population at risk M is the greatest integer not exceeding $N_0 n / (n - x)$. If this is an integer, then both $N_0 n / (n - x)$ and $n_0 n / (n - x) - 1$ maximizes the likelihood. Without loss of generality we use here $\hat{M} = N_0 n / (n - x)$. Observe that if \hat{M} is the MLE of M , then since $M = N_0 + [N_1\beta]$ we have that the MLE of β is

$$(\hat{M} - N_0) / N_1 \quad (6)$$

that is

$$\frac{x / N_1}{(n - x) / N_0} \quad (7)$$

which is the estimator of Greenwood and Yule [1]. Nevertheless, the construction here based on (5) allows for the derivation of some important properties of this estimator, for instance, Chapman [10] discussed the estimator \hat{M} , and since it is biased he suggested using instead

$$M^* = \frac{(N_0 + 1)(n + 1)}{n - x + 1} - 1$$

which has [11]

$$E[M^* - M] = \frac{(N_0 + 1)(n + 1)(M - N_0)!(M - n)!}{(M + 1)!(M - n - N_0 - 1)!} \quad (8)$$

From (6), an estimator of β is

$$\hat{\beta} = \frac{(N_0 + 1)x}{N_1(n - x + 1)} \quad (9)$$

with bias

$$E[\hat{\beta} - \beta] = \frac{(N_0 + 1)(n + 1)[N_1\beta]!(N_0 + [N_1\beta] - n)!}{N_1(N_0 + [N_1\beta] + 1)!([N_1\beta] - n - 1)!} \quad (10)$$

which is negligible for n small compared to N_1 and N_0 . The variance of M^* is approximately

$$\text{var}(M^*) = \frac{M^2 n N_0 (M - N_0) (M - n) (n + 1)^2 (N_0 + 1)^2}{(M - 1) (M + nN_0)^4}$$

from (6) $\text{var}(\hat{\beta}) = \text{var}(M^*)/N_1^2$, thus

$$\text{var}(\hat{\beta}) = \frac{M^2 n N_0 (M - N_0) (M - n) (n + 1)^2 (N_0 + 1)^2}{(M - 1) N_1^2 (M + nN_0)^4} \quad (11)$$

with $M = N_0 + [N_1\beta]$.

Confidence intervals for β can be obtained directly from those obtained for M . If (L, U) is a $1 - \alpha$ CI for M then from (6)

$$\left(\frac{L - N_0}{N_1}, \frac{U - N_0}{N_1} \right)$$

is a $1 - \alpha$ CI and β . Chapman [10] derived large sample CIs for M . Hypothesis testing for β is also straightforward.

2.2. The case of a random number of protected individuals

As it was mentioned earlier, it is natural to assume that a model 2 vaccine with $\text{VE} = 1 - \beta$ confers total immunity to an individual according to a Bernoulli random variable with probability of success $1 - \beta$. Thus, Y , the number of vaccinated unprotected in the population is a binomial random variable with parameters (N_1, β) .

An approximation of $E(\hat{\beta})$ and $\text{var}(\hat{\beta})$ can be done assuming that the bias in the estimation of the total number of susceptible given by (10) is negligible, thus

$$E[N_1\hat{\beta} + N_0 | Y] = Y + N_0$$

therefore $E[\hat{\beta} | Y] = Y/N_1$ and

$$\text{var}[E(\hat{\beta} | Y)] = \beta(1 - \beta)/N_1 \quad (12)$$

where $\hat{\beta}$ is given by (9). From (11)

$$\text{var}(\hat{\beta} | Y) = \frac{n N_0 Y (N_0 + Y - n) (n + 1)^2 (N_0 + 1)^2 (N_0 + Y)^2}{(N_0 + Y - 1) N_1^2 (N_0 + Y + nN_0)^4}$$

Since $\text{var}(\hat{\beta}) = \text{var}[E(\hat{\beta} | Y)] + E[\text{var}(\hat{\beta} | Y)]$, using (12) we have

$$\text{var}(\hat{\beta}) = \beta(1 - \beta)/N_1 + E[\text{var}(\hat{\beta} | Y)] \quad (13)$$

The second term in the right side of (13) can be approximated by substituting $E[Y] = N_1 \beta$, yielding

$$\text{var}(\hat{\beta}) = \beta(1 - \beta)/N_1 + \frac{n N_0 \beta (N_0 + N_1 \beta - n) (n + 1)^2 (N_0 + 1)^2 (N_0 + N_1 \beta)^2}{(N_0 + N_1 \beta - 1) N_1 (N_0 + N_1 \beta + n N_0)^4} \quad (14)$$

2.3. Subsampling

If only N'_0 and N'_1 unvaccinated and vaccinated individuals, respectively, are followed during the outbreak, then let $M' = n'_0 + [N'_1 \beta]$ be the ‘total’ group size and the results of the previous section still apply, either considering $[N'_1 \beta]$ fixed or random. Of course, since at least $N'_0 < N_0$ or $N'_1 < N_1$, the consequences of working with subsamples are an increase in the variance of the estimators of VE.

3. INFERENCE FOR LEAKY (MODEL 1) VACCINES

Wallenius [12] derived a distribution that has received very little attention in the literature. He called the sampling procedure ‘biased sampling’ and the resulting distribution the non-central hypergeometric distribution. Wallenius [12] derived this distribution as a need to characterize the non-null distribution against random sampling in the following context: assume that a lot containing N_0 low quality and N_1 high quality items is dichotomized according to some quality criterion among two purchasers A and B. From a lot of n items purchaser B may wish to test if the supplier has favoured purchaser A.

Wallenius [12] also showed that his distribution can be constructed with sampling without replacement from an urn containing Np balls of type 1 and $N - Np$ of type 0, with the modification that a selected ball of type i is drawn with probability p_i , $i = 0, 1$, and is returned to the urn with probability $1 - p_i$. In this sampling scheme, the total sample size n corresponds only to the number of extractions from the urn without considering ‘failures’.

Infections in a population of size $N_0 + N_1$ with a vaccinated fraction of size N_1 occur similarly to Wallenius’ sampling model [12], where sampling is equivalent to a threat of infection. If a vaccinated individual is selected then the person is effectively infected (extracted) with probability β_1 , whereas unvaccinated individuals are infected with probability β_0 . After x and y infections among the vaccinated and unvaccinated, respectively, the probability that the next infected is a vaccinated individual is

$$\frac{(N_1 - x)\beta_1/\beta_0}{(N_0 - y) + (N_1 - x)\beta_1/\beta_0}$$

The derivation of the probability mass function of Wallenius’ distribution is rather complicated. Wallenius [12] obtained the following formula (adapted here to our notation):

$$P(X = x; n, N_0, N_1, \beta) = \binom{N_1}{x} \binom{N_0}{n-x} \int_0^1 (1-t)^x (1-t^{c/\beta})^{n-x} dt \quad (15)$$

with $c = (N_1 - x + (N_0 - n + x)/\beta)^{-1}$ and $\beta = \beta_1/\beta_0$. Clearly, of $\beta_1 = \beta_0$ then (15) equals the hypergeometric distribution.

Estimation of β in (15) through maximum likelihood or via the method of moments is particularly difficult. Wallenius' interest was the calculation of probabilities and left much work to do regarding this distribution. It is possible to find the MLE of β in (15) by using numerical integration as well as maximization procedures. If $\hat{\beta}$ is the MLE of β in (15), then we can estimate $E[X]$ and $\text{var}[X]$ by plugging in $\hat{\beta}$ and the known values of n , N_0 and N_1 in (15) and thus the variance of the estimator for leaky vaccines (3) can be approximated. Thus

$$\text{var}(\widehat{\text{VE}}) = \text{var}\left(\frac{\ln(1 - x/N_1)}{\ln(1 - (n - x)/N_0)}\right)$$

Using the delta technique for function of random variables

$$\text{var}(\widehat{\text{VE}}) = \sigma^2 \left[\frac{1}{(N_1 - \mu) \ln(1 - (n - \mu)/N_0)} + \frac{\ln(1 - \mu/n)}{(N_0 - n + \mu) \ln(1 - (n - \mu)/N_0)} \right]^2 \quad (16)$$

where $E[X]$ and $\text{var}[X]$ are μ and σ^2 , respectively, which are substituted by their numerical estimates.

3.1. Approximations to Wallenius' model

Wallenius [12] also showed that when n is small compared to N_0 and N_1 the number of balls of type 1 in a sample of size n can be approximated with a binomial distribution with parameters n and $\phi = p\beta/(1 - p + p\beta)$ where $p = N_1/(N_0 + N_1)$. Under these conditions the MLE of ϕ is x/n and the MLE of β becomes

$$\beta^* = \frac{(1 - p)x}{p(n - x)} = \frac{N_0}{N_1} \frac{\hat{\phi}}{1 - \hat{\phi}} = \frac{x/N_1}{(n - x)/N_0}$$

which is the same as (7), the estimator of $\hat{\beta}$ for a model 2 vaccine. This implies that when the conditions for the binomial approximation are met, the estimators of VE for model 1 and 2 vaccines converge. Since the variance of the estimator of the odds ratio [13] is approximately

$$\text{var}(\hat{\phi}/(1 - \hat{\phi})) \approx \frac{\hat{\phi}}{n(1 - \hat{\phi})^3}$$

it follows that an approximation to the variance of the estimator β is

$$\text{var}(\beta^*) \approx \left(\frac{N_0}{N_1}\right)^2 \frac{\beta^* p [1 - p(1 - \beta^*)]^2}{n(1 - p)^3}$$

If (ϕ_L, ϕ_U) is an approximate $1 - \alpha$ confidence interval for ϕ , then

$$P\left(\frac{\phi_L(1 - p)}{(1 - \phi_L)p} < \beta < \frac{\phi_U(1 - p)}{(1 - \phi_U)p}\right) > 1 - \alpha$$

is an approximate $1 - \alpha$ CI for β . Wallenius [12] suggested two other approximations to (15) to consider the cases in which N_0 , N_1 and n are large and when N_0 , N_1 , n and x are large.

4. SIMULATIONS

In this section we simulate infections among a partially vaccinated population assuming random mixing, to test the estimates and their variances. We assume that the type of action of the vaccine

Table I. Simulations.

N_0	N_1	n	β	All/nothing			Leaky		
				$\hat{\beta}^*$	SE($\hat{\beta}$) [†]	SE($\hat{\beta}$) [‡]	$\tilde{\beta}^§$	SE($\tilde{\beta}$) [¶]	SE($\tilde{\beta}$)
100	100	20	0.1	0.0998	0.0734	0.0768	0.1080	0.0748	0.0818
100	100	20	0.05	0.0502	0.0500	0.0529	0.0539	0.0500	0.0538
150	150	20	0.1	0.1004	0.0730	0.0781	0.1067	0.0761	0.0818
150	200	20	0.05	0.0501	0.0435	0.0469	0.0530	0.0447	0.0479
150	200	30	0.05	0.0500	0.0364	0.0374	0.0518	0.0360	0.0374
200	200	20	0.2	0.1993	0.1118	0.1200	0.2145	0.1187	0.1300
200	250	30	0.1	0.1002	0.0556	0.0574	0.1050	0.0565	0.0600
300	400	30	0.2	0.1999	0.0842	0.0894	0.2099	0.0888	0.0948
300	400	20	0.1	0.1003	0.0655	0.0700	0.1068	0.0685	0.0748

* VE estimate using (9)

† Using (11), the true SE under random protection

‡ SE of the estimate $\hat{\beta}$ over 30 000 simulations

§ VE estimate using (3)

¶ SE obtained numerically from (15). First $E[X]$ and $\text{var}[X]$ are obtained numerically by plugging in $\tilde{\beta}$ in (15) and these two values are used in (16)

|| SE of the estimate $\tilde{\beta}$ over 30 000 simulations

is known. Several values for N_0 , N_1 , n and β are tested, with 30 000 simulations for each set of parameters (see Table I).

The simulations for all/nothing vaccines were obtained assuming that the vaccine fully protects every individual with probability $1 - \beta$, which is a more realistic assumption. In general it can be seen that the point estimates perform very well, and that $S^2(\hat{\beta})$, the sample variance of the estimates, is close to the expected. For leaky vaccines the variance of the estimates is not known, so we used the approximation (16) plugging in estimates of μ and σ^2 obtained by numerical solution for these two moments in (15). It can also be seen that $S^2(\tilde{\beta})$, the sample variance of the estimates, is fairly close to the approximated variance.

5. EXAMPLES

We apply the methods here to the data set from Muyinga, Burundi measles outbreak [7]. This data set consists of the observed attack rates in the three groups of individuals, grouped by age. The data are in Table II.

The goal of this section is to illustrate the methodology, so we analyse and obtain point estimates for each group separately. A methodology to construct an overall estimate is beyond the scope of this paper and it can be found elsewhere [14].

Tables III and IV show the VE estimates assuming an all/nothing and leaky effect, respectively. For the all/nothing case, the standard error of the proposed estimate is smaller than that of the current estimate, either for the fixed or random number protected, but the point estimates are not very different, except for the last group. For the leaky case, the VE estimate obtained by numerical maximization of Wallenius' distribution is not very different from the current estimate proposed by Haber *et al.* [5] as well as the standard errors.

Table II. Muyinga measles data set.

Age group (months)	Unvaccinated			Vaccinated		
	N_{k0}	x_{k0}	AR_{k0}	N_{k1}	x_{k1}	AR_{k1}
1 [9–15]	109	62	0.568	90	16	0.177
2 [16–36]	84	34	0.404	449	60	0.133
3 [37–60]	19	3	0.157	413	33	0.079

N_{ki} , size of k th group with vaccination status i ; x_{ki} , number of infected in k th group with vaccination status i ; AR_{ki} , attack rate in k th group with vaccination status i .

Table III. Estimation for all/nothing vaccines (model 2).

Group	\widehat{VE}^*	SE^\dagger	\widehat{VE}^\ddagger	SE^\S	SE^\P
1	0.687	0.075	0.689	0.056	0.075
2	0.669	0.059	0.675	0.051	0.056
3	0.494	0.280	0.600	0.132	0.134

* VE estimate using (1)

† Estimate of the standard error of \widehat{VE} using (2)

‡ VE estimate using (9)

§ Estimate of the standard error of \widehat{VE} assuming fixed protection as in (11), plugging in $\beta = 1 - \widehat{VE}$

¶ Estimate of the standard error of \widehat{VE} assuming random protection as in (14), plugging in $\beta = 1 - \widehat{VE}$.

Table IV. Estimation for leaky vaccines (model 1).

Group	\widehat{VE}^*	SE^\dagger	\widehat{VE}^\ddagger	$E[\widehat{X}]^\S$	$\text{var } \widehat{E}[X]^\S$	SE^\P
1	0.766	0.065	0.766	16.00	9.01	0.065
2	0.723	0.059	0.723	59.76	27.9	0.081
3	0.515	0.292	0.514	33.00	2.34	0.293

* VE estimate using (3)

† Estimate of the standard error of \widehat{VE} using (4)

‡ VE estimate using the MLE of β from numerical maximization of (15)

§ Obtained by plugging in the MLE of β in (15)

¶ Estimate of the standard error of \widehat{VE} using (16)

6. DISCUSSION

Some methods for VE estimation require an underlying model for the transmission of the disease. The approach followed here shows that the current point estimates perform very well without the assumptions required for those epidemic models. We have shown how to derive the MLEs for all/nothing and leaky vaccines which provides a strong framework for CIs and hypothesis testing.

Changes in the contact rate during an epidemic would be somewhat equivalent to changes in the rate of extraction of balls from the urn, which should not affect the distribution of the number of balls of a given type after conditioning on a fixed sample size, as long as the changes in contact rates are the same for vaccinated and unvaccinated individuals. Therefore, some assumptions of

the modelling approach can be relaxed, for instance, the estimates are robust to changes in contact and removal rates, perhaps accounting for a change in behaviour during the outbreak. Also, neither the time between contacts nor the removal rates need to be exponentially distributed.

For all/nothing vaccines, the random protection model is clearly more realistic than the fixed one, and should preferably be used. In particular for the Muyinga data set, the urn model approach allowed for an estimate with smaller standard error for all/nothing vaccines. The increase on the standard error by assuming random protection over fixed protection is an amount $\beta(1 - \beta)/N_1$.

The hypergeometric distribution for all/nothing vaccines as well as Wallenius' for leaky vaccines still hold in the presence of subsampling, with n being the size of the subsample, therefore final attack rate data are not necessary; alternatively, only a random fraction of the population can be analysed. The cost, of course, is in an increase of the variance of the estimates.

ACKNOWLEDGEMENTS

This work was supported partially by grant 438100-5-I26932-M from CONACYT (México). We would like to thank Ira Longini, Elizabeth Halloran, Michael Haber and two anonymous referees for comments and suggestions to improve this work.

REFERENCES

1. Greenwood M, Yule UG. The statistics of anti-typhoid and anticholera inoculations, and the interpretation of such statistics in general. *Proceedings of the Royal Society of Medicine* 1915; **8**(part 2):113–194.
2. O'Neill RT. On sample sizes to estimate the protective efficacy of a vaccine. *Statistics in Medicine* 1988; **7**:1279–1288.
3. Halloran ME, Haber M, Longini IM, Struchiner CJ. Direct and indirect effects in vaccine efficacy and effectiveness. *American Journal of Epidemiology* 1991; **133**:323–331.
4. Halloran ME, Haber M, Longini IM. Interpretation and estimation of vaccine efficacy under heterogeneity. *American Journal of Epidemiology* 1992; **136**:328–343.
5. Haber M, Longini IM, Halloran ME. Measures of the effects of vaccination in a randomly mixing population. *International Journal of Epidemiology* 1991; **20**:300–310.
6. Farrington CP. The measurement and interpretation of age-specific vaccine efficacy. *International Journal of Epidemiology* 1992; **21**:1014–1020.
7. Longini IM, Halloran ME, Haber M, Chen RT. Measuring vaccine efficacy from epidemics of acute infectious agents. *Statistics in Medicine* 1993; **12**:249–263.
8. Smith PG, Rodrigues LC, Fine PEM. Assessment of the protective efficacy of vaccines against common diseases using case-control and cohort studies. *International Journal of Epidemiology* 1984; **13**:87–93.
9. Becker NG. Estimation in models for the spread of infectious diseases. Proceedings of the XIth International Biometrics Conference, INRA, Versailles, 1982; pp. 145–151.
10. Chapman DG. Some properties of the hypergeometric distribution with application to zoological sample census. *University of California Publications in Statistics* 1951; **1**:131–159.
11. Johnson NK, Kotz S, Kemp AW. *Univariate Discrete Distributions* 2nd edn, Wiley: New York, 1992.
12. Wallenius KT. Biased sampling: the noncentral hypergeometric probability distribution. PhD thesis, Stanford University: Stanford, CA, 1963.
13. Casella G, Berger RL. *Statistical Inference*. Duxbury Press: California, 1990.
14. Haber M, Longini IM, Halloran ME. Estimation of vaccine efficacy in outbreaks of acute infectious diseases. *Statistics in Medicine* 1991; **10**:1573–1584.