URN MODELS AND VACCINE EFFICAY ESTIMATION

Carlos M. Hernández-Suárez
CGIC - Universidad de Colima, Gonzalo de Sandoval 444
Colima, Colima; 28045 México
Ph (331) 4-11-33, Fax (331) 2-75-81
e-mail : cmh2@cgic.ucol.mx

Carlos Castillo-Chavez
Biometrics Unit, Cornell University, Warren Hall 435, Ithaca, NY 14850
Phone: (607)-255-8103, e-mail : cc32@cornell.edu
Instituto de Investigaciones en Matemáticas Aplicadas y Sistemas
Universidad Nacional Autónoma de México, México City, APDO Postal 20-726, México
e-mail: castillo@leibniz.iimas.unam.mx
Institute for Mathematics and its Applications,University of Minnesota
400 Lind Hall, 207 Church St. SE, Minneapolis, MN 55455-0463, 612-624-6066, FAX 612-626-7370
Department of Mathematics, Howard University, Washington, DC 20059

URN MODELS AND VACCINE EFFICAY ESTIMATION

Abstract

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. It is shown that for all/nothing vaccines, the distribution of the number of infected vaccinated individuals conditioning in 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 this distribution follows a noncentral hypergeometric distribution. We derive the MLE of the efficacy of all/nothing vaccines and show that that of a leaky vaccine can be found by using standard numerical methods. 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 is 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 like tetanus.

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 usual estimate of vaccine efficacy (VE) has the form:

$$1 - \frac{RR_1}{RR_0}, \tag{1}$$

where $RR_0(RR_1)$ are the relative risks of infection among the unvaccinated (vaccinated) individuals. The first measure of vaccine efficacy was suggested by Greenwood and Yule[1] as

$$VE = 1 - \frac{AR_1}{AR_0}, \tag{2}$$

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 the estimate (2)

$$\frac{(1 - AR_0)}{n_0 AR_0} + \frac{(1 - AR_1)}{n_1 AR_1}. \tag{3}$$

If a vaccine confers total immunity to a fraction $1 - \beta$ of the population and leaves susceptible the remaining fraction then it is called an "all/nothing"

or model 2 vaccine. If every vaccinated individual reduces her/his chances of infection to β every time s/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[3,4], Haber et al[5], Farrington[6] and Longini[7]. We use the measure of VE defined by Haber et al[5]:

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

where $1 - \beta_1$ corresponds to the protection of the vaccine and β_0 to the natural susceptibility of unvaccinated individuals. Smith et al[8] showed that if the vaccine decreases the probability of a disease given exposure (model 1) then the efficacy as in (2) will depend on time after beginning of exposure to infection. Thus, for leaky vaccines Haber et al[5] suggested the estimate

$$VE = 1 - \frac{\ln(1 - AR_1)}{\ln(1 - AR_0)}, \tag{4}$$

who has approximate variance[9]

$$\frac{AR_1/(n_1 - x) + (1 - VE)^2(AR_0/(n_0 - n + x))}{(\ln(1 - AR_0))^2}. \tag{5}$$

The estimates (2) and (4) are usually 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 of 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 to 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 on the VE.

In the next sections we derive the statistical distribution of the the number

of infected among vaccinated or unvaccinated for model 1 and 2 vaccines, and use these to construct estimates of VE for model 1 and 2 vaccines, keeping the amount of assumptions to a minimum.

2 Inference for all/nothing (model 2) vaccines

Assume all non-vaccinated individuals are susceptible to the disease, then $\beta_0 = 1$ and $VE = 1 - \beta_1$. 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 susceptibles, and *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 $[n_1\beta_1]$ 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 $1-\beta_1$ 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_1). First we assume the number of vaccinated is a fixed quantity $[n_1\beta_1]$ 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:

$$P(X=x;n,n_0,\beta_1)=\frac{\binom{[n_1\beta_1]}{x}\binom{n_0}{n-x}}{\binom{n_0+[n_1\beta_1]}{n}}, \tag{6}$$

for $x=0,1,2,...,\min(n,n_1)$. Since n_0,n_1 and n are known, and x is observed, we can make inferences on β_1 through inferences on the total population at risk $M=n_0+[n_1\beta_1]$ 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 estimate (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 $\widehat{M} = n_0 n/(n-x)$. Observe that if $\widehat{M}$ is the MLE of M, then since $M = n_0 + [n_1 \beta_1]$ we have that the MLE of β_1 is

$$\widehat{\beta_1} = (\widehat{M} - n_0)/n_1, \tag{7}$$

that is

$$\widehat{\beta_1} = \frac{x/n_1}{(n-x)/n_0)}, \tag{8}$$

which is Greenwood and Yule[1] estimate. Nevertheless, the construction here based on (6) allows for the derivation of some important properties of this estimate, for instance, Chapman[10] discussed the estimate $\widehat{M}$,since it is biased, he suggested using instead

$$M^* = \frac{(n_0+1)(n+1)}{n-x+1} - 1,$$

who has

$$E[M^* - M] = \frac{(n_0+1)(n+1)(M-n_0)!\,(M-n)!}{(M+1)!\,(M-n-n_0-1)!}. \tag{9}$$

From (7) and (9), an estimate of β_1 is:

$$\beta_1^* = \frac{(n_0+1)\,x}{n_1\,(n-x+1)}, \tag{10}$$

with bias

$$E[\beta_1^* - \beta_1] = \frac{(n_0+1)(n+1)[n_1\beta_1]\,!\,(n_0+[n_1\beta_1]-n)\,!}{n_1\,(n_0+[n_1\beta_1]+1)\,!\,([n_1\beta_1]-n-1)\,!}, \tag{11}$$

which is negligible for n small compared to n_1 and n_0. The variance of $\widehat{M}$ is approximately[11]

$$Var(\widehat{M}) = M^2\left(\frac{M}{n_0 n} + 2\left(\frac{M}{n_0 n}\right)^2 + 6\left(\frac{M}{n_0 n}\right)^3\right),$$

from (7) $Var\,(\beta_1^*) = Var\,(M^*)\,/n_1^2$, thus

$$Var(\beta_1^*) = \frac{M^3}{n_0\,n\,n_1^2}\left(1 + 2\frac{M}{n_0\,n} + 6\left(\frac{M}{n_0\,n}\right)^2\right), \tag{12}$$

with $M = n_0 + [n_1\beta_1]$.

Confidence intervals for β_1 can be obtained directly from those obtained for M. If (L, U) is a 1-α CI for M then from (7)

$$\left(\frac{L - n_0}{n_1}, \frac{U - n_0}{n_1}\right)$$

is a 1-α CI for β. Chapman[10] derived large sample CI's for M. Testing hypothesis on β_1 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 VE $= 1 - \beta_1$ confers total immunity to an individual according to a Bernoulli random variable with probability of success $1 - \beta_1$. Thus, Y, the number of vaccinated unprotected in the population is a Binomial random variable with parameters (n_1, β_1).

An approximation to $E(\widehat{\beta_1})$ and $Var(\widehat{\beta_1})$ can be done assuming that the

bias in the estimation of the total number of susceptible given by (11) is negligible, thus

$$E\left[n_1\widehat{\beta_1} + n_0|Y\right] = Y + n_0,$$

therefore $E[\widehat{\beta_1}|Y] = Y/n_1$ and

$$Var\left[E(\beta_1^*|Y)\right] = \beta_1(1-\beta_1)/n_1, \tag{13}$$

and from (12)

$$Var(\beta_1^*|Y) = \frac{(n_0+Y)^3}{n_0\, n\, n_1^2}\left(1 + 2\frac{n_0+Y}{n_0\, n} + 6\left(\frac{n_0+Y}{n_0\, n}\right)^2\right). \tag{14}$$

Since $Var(\beta_1^*) = Var[E(\beta_1^*|Y)] + E[Var(\beta_1^*|Y)]$, using (13) we have

$$Var(\beta_1^*) = \beta_1(1-\beta_1)/n_1 + E[Var(\beta_1^*|Y)]. \tag{15}$$

We give an expression for $E[Var(\beta_1^*|Y)]$ in the appendix.

2.3 Subsampling

If only $n_0^{'}$ and $n_1^{'}$ unvaccinated and vaccinated individuals respectively are followed during the outbreak, then let $M' = n_0^{'} + [n_1^{'}\beta_1]$ be the "total" group size and the results of the previous section still applies, either considering $[n_1^{'}\beta_1]$ fix 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 estimates 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 noncentral hypergeometric distribution. (The extended hypergeometric distribution was mistakenly referred as the noncentral hypergeometric distribution by Ewell[13] while constructing confidence intervals for the vaccine efficacy). 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 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 favored purchaser A.

Wallenius[12] also showed that the noncentral hypergeometric 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 $1 - p_i$. It is important to emphasize that in this sampling scheme, the total sample size n corresponds only to the number of extractions from the urn without considering "failures".

Notice that infections in a population of size $n_0 + n_1$ with a vaccinated fraction of size n_1 occurs similarly to Wallenius' sampling model[12], were sampling is equivalent to a threat of infection and if a vaccinated individual is selected this 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 infection is a vaccinated individual is

$$\frac{(n_1 - x)\beta_1/\beta_0}{(n_1 - y) + (n_1 - x)\beta_1/\beta_0}.$$

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

$$P(X = x; n, n_0, n_1, \beta) = \binom{n_1}{x}\binom{n_0}{n-x}\int_0^1 (1-t^c)^x \left(1-t^{c/\beta}\right)^{n-x} dt, \quad (16)$$

with $c = (n_1 - x + (n_0 - n + x)/\beta)^{-1}$ and $\beta = \beta_1/\beta_0$. Clearly, if $\beta_1 = \beta_0$ then (16) equals the hypergeometric distribution.

Estimation of β in (16) through maximum likelihood or via the method of moments is particularly difficult. Wallenius[12,] 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 (16) by using numerical integration as well as maximization procedures. If $\widehat{\beta}^*$ is the MLE of β in (16), then we can estimate $E[X]$ and $Var[X]$ by plugging in $\widehat{\beta}^*$ and the known values of n, n_0 and n_1 in (16) and thus the variance of the estimate for leaky vaccines (4) can be approximated. Thus,

$$Var(\widehat{VE}) = Var\left(\frac{\log(1 - x/n_1)}{\log(1 - (n-x)/n_0)}\right).$$

Using the delta technique for function of random variables

$$Var(\widehat{VE}) = \sigma^2\Big[\frac{1}{(n_1-\mu)\log(1-(n-\mu)/n_0)} + \frac{\log(1-\mu/n)}{(n_0-n+\mu)\log(1-(n-\mu)/n_0))}\Big]^2 \tag{17}$$

where μ and σ^2 are substituted by their numerical estimates.

3.1 Approximations to Wallenius' model

It is possible to show[12] 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 this conditions the MLE estimate of ϕ is x/n and the MLE estimate of β becomes

$$\widehat{\beta} = \frac{(1-p)\,x}{p\,(n-x)} = \frac{n_0}{n_1}\frac{\widehat{\phi}}{1-\widehat{\phi}} = \frac{x/n_1}{(n-x)/n_0},$$

which is the same as (8), the estimate of $\widehat{\beta}$ for a model 2 vaccine. This implies that when the conditions for the binomial approximation are met,

the estimate of VE for model 1 and 2 vaccines converge. Since the variance of the estimate of the odds ratio[15] is approximately

$$Var\left(\widehat{\phi}(1-\widehat{\phi})\right) \approx \frac{\widehat{\phi}}{n\left(1-\widehat{\phi}\right)^3},$$

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

$$Var(\widehat{\beta}) \approx \left(\frac{n_0}{n_1}\right)^2 \frac{\widehat{\beta}\, p\left[1-p\left(1-\widehat{\beta}\right)\right]^2}{n(1-p)^3}.$$

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-α CI for β. Wallenius[12] suggested two other approximations to (16) 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 Inference for non-random mixing

It is possible to estimate VE separately for groups of individuals as long as the appropriate distribution (hypergeometric or noncentral hypergeometric) still applies for every group. This is only possible when the probability

that the next infection in a group is a vaccinated or unvaccinated individual depends only on the relative proportion of susceptiles on each group. This condition is met if the population can be divided in groups where a contact of an individual of group i with an individual of group j occurs at a rate $\lambda_i q_{ij}$. Therefore, a group can be formed with individuals that have the same contact rate, mix randomly among them and have the same probability of contact with every other group. Observe that no knowledge is required about the mixing patterns between groups or contact rates.

Once the vaccine efficacy is estimated for every group, an overall estimate can be constructed as

$$VE^* = 1 - \sum_{u=1}^{k} w_u \, \widehat{\theta}_u \tag{18}$$

with $\widehat{\theta}_u$ being the estimate VE for group u and w_u is the weight given to the estimate in group u. A good choice for the w_u would be

$$w_u = \frac{\sigma_i^{-2}}{\sum_i \sigma_i^{-2}} \tag{19}$$

the reciprocal of the normalized variance in group u, as in Haber et al[14], (see also Casella and Berger[15], p.338), although it must be pointed out that the estimates are biased. In practice we would substitute an estimate of this

variance. The estimated variance of VE^* in (18) becomes

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

5 Examples

We apply the methods here to the data set from Muyinga, Burundi measles outbreak[7]. This data set consists on the observed attack rates in three groups of individuals, grouped by age. The data is in Table 1.

Table 1 Approx. here

Table 2 shows the VE estimates assuming an all/nothing effect. The VE estimate using (10) is markedly different from the estimated VE using (1) for age group 3. If the amount of vaccinated unprotected is a fixed value $[n_1\beta_1]$ then the weighted VE estimate is VE = 0.673, with s.e. = 0.079. If we assume a random number of unprotected individuals the weighted estimate of VE is VE = 0.667 with s.e. = 0.144, which almost doubles the standard error for fixed protection.

Table 2 Approx. here

Table 3 shows the VE estimates under the assumption that the vaccine is leaky. It can be seen that the usual estimate (2) is similar to the MLE estimate of VE by maximizing (16). Standard errors are similar too. The standard errors for the VE in column 7 were obtained by calculating numerically the MLE of β in (16) and then computing $\widehat{E[X]}$ and $\widehat{Var[X]}$. This values were used as μ and σ^2 in (17). *Mathematica*[16] was used for numerical integration.

Table 3 approx. here

The $VE^{(*)}$ and $VE^{(**)}$ both estimate the vaccine efficacy in a stratified population as long as the homogeneous mixing assumption is valid for every group. We can construct a weighted estimate of VE as in (18) using weights according to (19), with σ_i =standard error for group i. The last column of Table 3 gives the weight using s.e. (††). In this way the estimate of vaccine efficacy is 0.7423, with standard error 0.050.

6 Discussion

Most of published work on VE evaluation deal require an underlying transmission model. The model-free approach followed here shows that the current

point estimates perform very well. We have shown how to derive the MLE's for all/nothing and leaky vaccines which provides a strong framework for CI's and hypothesis testing since these kind of estimates are asymptotically normal.

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. Therefore, some assumptions of the modeling approach can be relaxed, for instance, the estimates are robust to changes in contact and removal rates for group k from λ_k and μ_k to $\lambda_k(t)$ and $\mu_k(t)$ respectively, perhaps accounting for a change in behavior during the outbreak. Also, neither the time between contacts nor the removal rates need to be exponentially distributed.

In particular for the Muyinga data set, the urn model approach allowed for an estimate with smaller bias for all/nothing vaccines.Thus, the standard error of the estimate (1) calculated from (2) was larger than that of (10) when we assumed a fixed number of vaccinated protected. When assumed a random number of vaccinated protected, the resulting standard errors were smaller than those of estimate (1) except for group 2. The hypergeometric

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

References

1. Greenwood M. and Yule U.G. "The statistics of anti-typhoid and anti-cholera inoculations, and the interpretation of such statistics in general", *Proceedings of the Royal Society of Medicine* **8**,(part 2),113-94 (1915).

2. O'Neill, R.T. "On sample sizes to estimate the protective efficacy of a vaccine", *Statistics in Medicine,* **7**,1279-1288 (1988).

3. Halloran, M.E., Haber,M, Longini,I.M. and Struchiner,C.J. "Direct and indirect effects in vaccine efficacy and effectiveness", *American Journal of Epidemiology,* **133**,323-331 (1991).

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

5. Haber, M, Longini, I.M. and Halloran, M.E. "Measures of the effects of vaccination in a randomly mixing population", *International Journal of Epidemiology,* **20**,300-310 (1991).

6. Farrington, C.P. "The measurement and interpretation of age-specific vaccine efficacy", *International Journal of Epidemiology,* **21**, 1014-1020 (1992).

7. Longini, I.M, Halloran, M.E., Haber, M. and Chen, R.T. "Measuring vaccine efficacy from epidemics of acute infectious agents", *Statistics in Medicine,* **12**, 249-263 (1993).

8. Smith,P.G., Rodrigues,L.C. and Fine, P.E.M. "Assessment of the protective efficacy of vaccines against common diseases using case-control and cohort studies", *International Journal of Epidemiology,* **13**, 87-93 (1984).

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

10. Chapman, D.G. "Some properties of the hypergeometric distribution with application to zoological sample census", University of California Publications in Statistics, **1**,131-159 (1951).

11. Johnson, N.K., Kotz, S. and Kemp, A.W. *Univariate discrete distributions,* 2nd edn, Wiley: New York, 1992.

12. Wallenius, K.T. "Biased Sampling: The Noncentral Hypergeometric Probability Distribution", Ph. D. thesis, Stanford, CA: Stanford University (1963).

13. Ewell, M. "Comparing methods for calculating confidence intervals for vaccine efficacy", *Statistics in Medicine,* **15**, 21-22 (1996).

14. Haber, M, Longini, I.M. and Halloran, M.E. "Estimation of vaccine efficacy in outbreaks of acute infectious diseases", *Statistics in Medicine,***10**, 1573-1584 (1991).

15. Casella, G. and Berger, R.L. *Statistical Inference,* Duxbury Press, California, 1990.

16. Wolfram, S. *The Mathematica Book*, 3rd edn, Wolfram Media, Cambridge University Press, 1966.

Appendix

Expanding $Var(\beta_1^*|Y)$ in (14) we have

$$\begin{aligned}
&Var(\beta_1^*|Y) = Y^5\left(\tfrac{6}{n^3n_0^3n_1^2}\right) + Y^4\left(\tfrac{30}{n^3n_0^2n_1^2} + \tfrac{2}{n^2n_0^2n_1^2}\right)\\
&+Y^3\left(\tfrac{60}{n^3n_0n_1^2} + \tfrac{8}{n^2n_0n_1^2} + \tfrac{1}{n^2n_0n_1^2}\right) + Y^2\left(\tfrac{60}{n^3n_1^2} + \tfrac{12}{n^2n_1^2} + \tfrac{60}{nn_1^2}\right)\\
&+Y\left(\tfrac{30\,n_0}{n^3n_1^2} + \tfrac{8\,n_0}{n^2n_1^2} + \tfrac{3\,n_0}{nn_1^2}\right) + \tfrac{n_0^2}{n\,n_1^2}\left(1 + \tfrac{2}{n} + \tfrac{6}{n^2}\right).
\end{aligned}$$

or

$$\begin{aligned}
Var(\beta_1^*|Y) \;&= K_5Y^5 + K_4Y^4 + K_3Y^3 + K_2Y^2\\
&\quad +K_1Y + K_0.
\end{aligned}$$

Therefore

$$\begin{aligned}
E[Var(\beta_1^*|Y)] \;&= K_5E[Y^5] + K_4E[Y^4] + K_3E[Y^3] + K_2E[Y^2]\\
&\quad +K_1E[Y] + K_0.
\end{aligned}$$

Since Y is Binomial(n_1, β) we have:

$$\begin{aligned}
E[Y] =\;& n_1\beta_1\\
E[Y^2] =\;& n_1\beta_1 + n_1(n_1-1)\beta_1^2\\
E[Y^3] =\;& n_1\beta_1 + 3\,n_1(n_1-1)\beta_1^2 + n_1(n_1-1)(n_1-2)\beta_1^3\\
E[Y^4] =\;& n_1\beta_1 + 7\,n_1(n_1-1)\beta_1^2 + 6n_1(n_1-1)(n_1-2)\beta_1^3\\
&+n_1(n_1-1)(n_1-2)(n_1-3)\beta_1^4\\
E[Y^5] =\;& n_1\beta_1 + 15\,n_1(n_1-1)\beta_1^2 + 25n_1(n_1-1)(n_1-2)\beta_1^3\\
&+10n_1(n_1-1)(n_1-2)(n_1-3)\beta_1^4\\
&+n_1(n_1-1)(n_1-2)(n_1-3)(n_1-4)\beta_1^5.
\end{aligned}$$

TABLE CAPTIONS

Table 1. Muyinga measles data set

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 2. Estimation for all/nothing vaccines (Model 2)

(*) VE estimate using (1); (**) standard error of $\widehat{VE}$ using (2); (†) VE estimate using (10); (††) standard error assuming fixed protection as in (12); (‡) standard error assuming random protection as in (15); (§) weights calculated assuming a fixed protection; (§§) weights calculated assuming random protection.

Table 3. Estimation for leaky vaccines (Model 1)

(*) using (4); (†) standard error of $\widehat{VE}$ using(5); (**) using the MLE of β from numerical maximization of (16);(‡) plugging in the MLE of β; (††)standard error of $\widehat{VE}$ using (17); (§) weights calculated using (22).

Table 1. Muyinga measles data set

	Unvaccinated			Vaccinated		
Age group (months)	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

Table 2. Estimation for all/nothing vaccines (Model 2)

Group	$\widehat{VE}^{*}$	$s.e.^{(**)}$	$\widehat{VE}^{\dagger}$	$s.e.^{(\dagger\dagger)}$	$s.e.^{(\ddagger)}$	$w_u^{\S}$	$w_u^{\S\S}$
1	0.687	0.241	0.689	0.196	0.219	0.163	0.431
2	0.669	0.178	0.675	0.089	0.235	0.777	0.375
3	0.494	0.555	0.600	0.324	0.327	0.059	0.193

Table 3. Estimation for leaky vaccines (Model 1)

Group	$\widehat{VE}^{*}$	$s.e.^{(\dagger)}$	$\widehat{VE}^{**}$	$\widehat{E[X]}^{(\ddagger)}$	$\widehat{Var E[X]}^{(\ddagger)}$	$s.e.^{(\dagger\dagger)}$	$w_u^{\S}$
1	0.766	0.065	0.766	16.00	9.01	0.065	0.585
2	0.723	0.059	0.723	59.76	27.9	0.081	0.384
3	0.515	0.292	0.514	33.00	2.34	0.293	0.029