



Published in final edited form as:

*Biometrics*. 2013 March ; 69(1): 235–244. doi:10.1111/j.1541-0420.2012.01826.x.

## ESTIMATING STRAIN-SPECIFIC AND OVERALL EFFICACY OF POLYVALENT VACCINES AGAINST RECURRENT PATHOGENS FROM A CROSS-SECTIONAL STUDY

**Kari Auranen**\*

Department of Vaccination and Immune Protection, National Institute for Health and Welfare, Finland

**Hanna Rinta-Kokko**, and

Department of Vaccination and Immune Protection, National Institute for Health and Welfare, Finland

**M. Elizabeth Halloran**

Center for Statistics and Quantitative Infectious Diseases, Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center and Department of Biostatistics, University of Washington

### Summary

Evaluating vaccine efficacy for protection against colonisation with bacterial pathogens is an area of growing interest. Colonisation of the nasopharynx is an asymptomatic carrier state responsible for person-to-person transmission. It differs from most clinical outcomes in that it is common, recurrent and observed only in its prevalent state. To estimate rates of acquisition and clearance of colonisation requires repeated active sampling of the same individuals over time, an expensive and invasive undertaking. Motivated by feasibility constraints in efficacy trials with colonisation endpoints, investigators have been estimating vaccine efficacy from cross-sectional studies without principled methods. We present two examples of vaccine studies estimating vaccine efficacy from cross-sectional data on nasopharyngeal colonisation by *Streptococcus pneumoniae* (pneumococcus). This study presents a framework for defining and estimating strain-specific and overall vaccine efficacy for susceptibility to acquisition of colonisation ( $VE_{acq}$ ) when there is a large number of strains with mutual interactions and recurrent dynamics of colonisation. We develop estimators based on one observation of the current status per study subject, evaluate their robustness, and re-analyse the two vaccine trials. Methodologically, the proposed estimators are closely related to case-control studies with prevalent cases, with appropriate consideration of the at-risk time in choosing the controls.

### Keywords

Case-control studies; Multiple colonisation; Pneumococcus; Polyvalent vaccine; Recurrent infection; Risk-set sampling; Strain interaction; Vaccine efficacy

---

\*kari.auranen@thl.fi.

8. Supplementary Materials

Web Appendices A, B and C referenced in Sections 3, 4, 5 and 6 are available under the Paper information link at the Biometrics website <http://www.biometrics.tibs.org/>.

## 1. Introduction

Vaccine efficacy (VE) is generally defined as a relative reduction in some measure of risk in the vaccinated group compared to the unvaccinated group (Halloran, Longini, and Struchiner, 2010). When a vaccine is polyvalent, that is, it contains antigens against two or more, but not necessarily all, related strains of a microorganism, assessing both the strain-specific efficacies and the overall (aggregate) efficacy against all strains is of interest. Some bacterial pathogens asymptotically colonise the nasopharyngeal passages, called the carrier state. Colonisation with any strain may be cleared and recur repeatedly. Colonisation only rarely leads to manifest disease, such as bacteremia or meningitis, but it is a prerequisite for disease and the only source of transmission. Thus, the effect of vaccination on colonisation is of interest. Examples of such bacterial pathogens with polyvalent vaccines either available or under development include *Streptococcus pneumoniae* (pneumococcus), *Neisseria meningitidis* (Tan, Carlone, and Borrow, 2010), and group B streptococci (Edwards, 2008). Current formulations of pneumococcal conjugate vaccine contain antigens of 7, 10 or 13 of the known 93 serotypes (MMWR, 2010).

A straightforward approach to define and estimate vaccine efficacy for susceptibility based on the hazard rate of acquiring colonisation ( $VE_{acq}$ ) for individual strains would be to treat colonisation by each strain as separate and independent outcomes. To estimate rates of acquisition and clearance of colonisation requires repeated active sampling of the same individuals over time, an expensive and invasive undertaking. Thus, investigators have turned to using cross-sectional data to estimate vaccine efficacy without, however, explicitly defining the estimand of interest and its underlying assumptions (Rinta-Kokko et al. 2009).

We present two randomised controlled trials of polyvalent pneumococcal vaccines in which assessment of serotype-specific and overall vaccine efficacy is based on cross-sectional data (Table 1). In a vaccine study in South Africa (Mbelle et al., 1999), 500 infants were randomised to either a 9-valent pneumococcal conjugate vaccine or control vaccine and given 3 doses at 6, 10, and 14 weeks of age. Serotype-specific colonisation was measured once at the age of 9 months. Samples were available from 481 children. In a vaccine trial in Finland in 1995–1999 (Kilpi et al., 2001), 2497 children were randomised to a 7-valent pneumococcal conjugate vaccine or control and given 3 doses at 2, 4 and 6 months of age, and a booster at 12 months of age. Serotype-specific colonisation was measured before the booster at 12 months of age and approximately 6 months after the booster at 18 months of age. Here we analyse measurements made at least 6 months after the booster (2403 children).

Rinta-Kokko et al. (2009) proposed to estimate  $VE_{acq}$  based on one cross-sectional study under stationary conditions using simple ratios of the odds of vaccination in those colonised with the types targeted by the vaccine versus those not colonised. However, that analysis was not adjusted for the possible difference in the times spent susceptible by the vaccinees and controls. Colonisation is usually common and recurrent even on the strain level. In addition, the time spent at-risk for acquiring a new colonisation is affected by between-strain interactions induced by intra-species competition (Lipsitch et al., 2000; Melegaro et al., 2007). Any reduction in colonisation within a vaccinated host by pneumococcal strains targeted by the vaccine is thus counterbalanced by an increase in non-vaccine type colonisation (e.g. Mbelle et al., 1999), a phenomenon we call within-host replacement. For these reasons, the vaccine efficacy estimands may have to be conditioned on the appropriate states of colonisation in which an individual is susceptible to colonisation by the strain(s) targeted by the vaccine. Without such adjustment, the biological protective efficacy on the causal pathway and the estimated (outcome) efficacy may be different, confounding the

interpretation and comparability of estimates of vaccine efficacy across different epidemiologic settings (cf. Struchiner et al., 1994).

Here we present a framework for defining vaccine efficacy for multiple interacting strains which can be acquired and carried repeatedly. We develop two new estimands of  $VE_{\text{acq}}$  in the multi-type setting. We propose estimators of  $VE_{\text{acq}}$  for individual strains as well as for any collection of select (target) strains based solely on one measurement per study subject. We generalise the model to allow estimation of combined efficacy against acquisition and duration of colonisation as a more general estimand within a broader class of vaccine efficacy on colonisation. We present simulations and discuss robustness of estimation to departures from model assumptions.

## 2. Models of colonisation and vaccine effects

Figure 1A depicts two alternative models (*A* and *B*) to describe recurrent colonisation and competition among multiple strains of a pathogen within an individual (Lipsitch, 1997; Melegaro et al. 2007; Auranen et al., 2010).

In model *A* for two strains, an individual can be either *susceptible* (i.e., uncolonised in state 0) or *colonised* with either of the two strains (in states 1 or 2). In addition, the individual may be simultaneously colonised by both strains (co-colonised state 1&2). Figure 1A presents the model only for two strains, but we consider it for an arbitrary number of distinguishable strains. However, in this paper we do not allow simultaneous colonisation with more than two strains, which is supported by the rarity of such multiple colonisation by pneumococci (Kaltoft et al., 2008). For  $n$  strains, the set of  $n(n+1)/2+1$  possible states of colonisation is denoted by  $\mathcal{S} = \{0\} \cup \mathcal{S} \cup \{(1, 2), (1, 3), \dots, (1, n), (2, 3), \dots, (n-1, n)\}$  where  $\mathcal{S} = \{1, \dots, n\}$  is the index set of the strains.

There are altogether  $2n^2$  transition hazards between the directed pairs of states in  $\mathcal{S}$ . For clarity, we index the individual states explicitly by strain indices, i.e.,  $q_{i,j}$  denotes the hazard of transition from state  $i$  to state  $j$ ,  $i, j \in \mathcal{S} \cup \{0\}$ ,  $i \neq j$ , and  $q_{i,ij}$  and  $q_{ij,i}$  denote the hazards of the two reverse transitions between states  $i$  and  $(i, j)$ ,  $i, j \in \mathcal{S}$ ,  $i \neq j$ . Occasionally, it is more convenient to use an alternative naming convention in which the states are enumerated as  $\{[1], \dots, [n(n+1)/2 + 1]\}$ . Then  $q_{[h][k]}$  denotes the hazard of transition from state  $[h]$  to state  $[k]$  ( $[h], [k] \in \mathcal{S}$ ,  $[h] \neq [k]$ ).

We assume the hazards of transition in the unvaccinated subjects (“*C*” for controls) under model *A* fulfil two conditions:

$$q_{j,ij}^C/q_{i,ij}^C = q_{0,i}^C/q_{0,j}^C \text{ for all } i, j \in \mathcal{S}, i \neq j, \quad (\text{A1})$$

$$q_{i,0}^C = q_{i,j}^C \equiv \mu \text{ for all } i, j \in \mathcal{S}, i \neq j. \quad (\text{A2})$$

For any two strains, assumption (A1) states the ratio of their two hazards of acquisition to the doubly-colonised state ( $j$  to  $(i, j)$  vs.  $i$  to  $(i, j)$ ) equals the ratio of the respective hazards of colonisation from state 0 (0 to  $i$  vs. 0 to  $j$ ). Competition between strains is quantified in terms of relative hazards  $q_{j,ij}^C/q_{i,ij}^C$ , values  $< 1$  indicating reduced acquisition of strain  $i$  due to the currently colonising strain  $j$ . Consequently, we refer to condition (A1) as symmetry in competition because it implies  $q_{j,ij}^C/q_{0,i}^C = q_{i,ij}^C/q_{0,j}^C$  for any given pair of strains  $i$  and  $j$ .

Conditions (A1) and (A2) still allow each strain  $i$  to have its own hazard of colonisation  $q_{0,i}$ . According to (A2), however, the hazards of clearance from the singly-colonised states to state 0 and the hazards of clearance from doubly-colonised states to singly-colonised states are equal for all strains and denoted by  $\mu$ . In Sections 5 and 6 we relax this assumption.

### Illustration

To illustrate these concepts, assume there were four strains, with the vaccine containing antigens of two strains, called the vaccine strains. The other two are *non-vaccine strains*. Denote the two vaccine strains 1 and 2, and the non-vaccine strains 3 and 4. Let the hazards of colonisation in uncolonised, unvaccinated individuals be  $q_{0,1} = 0.07$ ,  $q_{0,2} = 0.06$ ,  $q_{0,3} = 0.05$ , and  $q_{0,4} = 0.04$  per time unit. Assume the hazard of clearance ( $\mu$ ) is 0.17 per time unit. Let the relative hazards for acquisition of double carriage ( $q_{j,ij}/q_{0,i}$ ) be 0.4 for all  $i, j$ , according to eq. (A1). The value  $0.4 < 1$  indicates considerable competition against double colonisation. For example, if a person is already colonised with strain 1, 3, or 4, the hazard of becoming doubly colonised with strain 2 would be  $0.06 \cdot 0.4 = 0.024$ .

### Model B

Figure 1B presents model  $B$  for two strains. Again, an individual can be either susceptible or colonised with either of the two strains, but there is no co-colonisation. In a model for  $n$  strains, the conditions corresponding to (A1) and (A2) are:

$$q_{j,i}^C/q_{i,j}^C = q_{0,i}^C/q_{0,j}^C, \quad i, j \in S, \quad i \neq j, \quad (\text{B1})$$

$$q_{i,0}^C = \mu, \quad i \in S. \quad (\text{B2})$$

Under model  $B$ , competition between strains is described in terms of the ratios  $q_{j,i}^C/q_{0,i}^C$  of the reverse transition hazards between the singly colonised states, values  $< 1$  again indicating competition. Although many results derived in this paper are applicable for any number of strains also under model  $B$ , the main use of this model is with three states only, obtained by aggregating states of model  $A$  (see Section 5).

### The vaccine model

We assume the direct biological effect of vaccination acts on susceptibility to acquisition of colonisation, the primary estimand of interest thus being  $\text{VE}_{\text{acq}}$ . Under model  $A$  a strain-specific efficacy estimand  $\text{VE}_{\text{acq},i}$  is defined as the relative reduction in the hazard of acquisition of strain  $i$  if vaccinated ( $T$ ) compared to if receiving control ( $C$ ):

$$\text{VE}_{\text{acq},i} = 1 - q_{0,i}^T/q_{0,i}^C = 1 - q_{j,ij}^T/q_{j,ij}^C, \quad i, j \in S, \quad i \neq j. \quad (\text{A3})$$

$\text{VE}_{\text{acq},i}$  is defined separately for each state from which a transition is possible to state  $i$  or ( $i, j$ ), i.e., for all states in which an individual is susceptible to acquisition of strain  $i$ . The definition assumes no interaction between the current state of colonisation and vaccine efficacy for strain  $i$ .

### Illustration, cont'd

In the four strain example above, let  $\text{VE}_{\text{acq},1} = 0.7$  and  $\text{VE}_{\text{acq},2} = 0.4$ . Under (A3), the efficacy against colonisation by strain 1 is the same whether an individual is uncolonised or

colonised with strain 2, 3, or 4. The hazard of an uncolonised vaccinated person being colonised with strain 1 would be  $(1 - 0.7)0.07 = 0.21$  (per time unit). However, due to the competition described above, the hazard of a colonised vaccinated person being colonised with strain 1 would be  $(1 - 0.7)0.4 \cdot 0.07 = 0.064$ . The same arguments hold for strain 2. The hazards of the non-vaccine strains 3 and 4 are unaffected by vaccination.

Initially we assume that vaccination does not affect the hazard of clearance of colonisation:

$$q_{i,0}^T = q_{i,j}^T = \mu, \quad i, j \in S, i \neq j. \quad (\text{A4})$$

Under model  $B$ , the analogous vaccine efficacy estimand is

$$\text{VE}_{\text{acq},i} = 1 - q_{0,i}^T / q_{0,i}^C = 1 - q_{ji}^T / q_{ji}^C, \quad i, j \in S, i \neq j, \quad (\text{B3})$$

and the assumption about the hazard of clearance is (B4)  $q_{i,0}^T = \mu, i \in S$ . It is straightforward to show that with the assumed vaccine models, (A3)–(A4) or (B3)–(B4), the respective conditions (A1)–(A2) or (B1)–(B2) hold for the vaccinated individuals as well. In the sequel, we drop “acq” in the notation for  $\text{VE}_{\text{acq},i}$  if not required for clarity.

### 3. Overall vaccine efficacy for multiple strains

We denote the index sets of the non-vaccine and vaccine strains by  $\bar{V}$  and  $V$ , so  $S = \bar{V} \cup V$ . We also define the set  $\bar{V}_0 = \bar{V} \cup \{0\}$  as the index set of those (single) states against which the vaccine does not confer direct biological protection. In particular, when conditioning on the current state of colonisation, assuming no cross-reactivity of immunity between strains, the true direct biological vaccine efficacy against a non-vaccine strain is zero, i.e.,  $\text{VE}_i = 0$  for  $i \in \bar{V}$ . In contrast, the efficacy against any vaccine strain is *a priori* unknown. We denote the index set of these target strain(s) in a vaccine study by  $W \subseteq V$ . Typically  $W = \{i\}$  separately for each  $i \in V$  (for strain-specific estimates for vaccine efficacy) or  $W = V$  (for overall efficacy against all vaccine-strains). However, there may be interest in a proper subset of (vaccine) strains as the target, e.g., “new” strains in a vaccine with a larger valency.

#### Partition of the states of colonisation

To define the estimands of overall efficacy, we need to specify a partition of the *states of colonisation* into three disjoint sets (Figure 2). The sets correspond to (a) states in which none of the vaccine-strains is involved, denoted  $\bar{v}_0$ ; (b) states in which any one of the target strains of the vaccine is involved either alone or with any of the non-vaccine strains, denoted  $\mathcal{W}$ ; and (c) the rest  $\mathcal{R} = S \setminus (\mathcal{W} \cup \bar{v}_0)$ . The sets  $\mathcal{W}$  and  $\mathcal{R}$  generally depend on the choice of target strains  $W$  against which vaccine efficacy is considered. For  $W = V$ ,  $\mathcal{R}$  contains only the doubly-colonised states with two vaccine-strains (Figure 2A). These states are not included in  $\mathcal{W}$  because our definition of vaccine efficacy does not involve acquisition of secondary vaccine strains (see below). For strain-specific efficacy,  $\mathcal{R}$  also includes all states with non-target vaccine-strains (Figure 2B).

#### Overall vaccine efficacy

Generalising the strain-specific estimands in (A3) and (A4), we define overall (aggregate) vaccine efficacy against all strains in a target set  $W$  as the relative reduction in their overall hazard of acquisition. To make this definition precise in a multi-state setting, we specify a distribution for the time spent at risk to acquire the target strains among the non-vaccine

states. We consider two estimands based on transitions either from state 0 only, or alternatively from any of the non-vaccine states  $\bar{V}_0$ .

The first estimand specifies the overall vaccine efficacy against a given collection of target strains as the relative reduction in their overall hazard of acquisition from state 0. The target states  $W$  then include all states to which there is a direct transition (“arrow”) from state 0 and the overall hazard is simply the sum of the hazards of colonisation  $q_{0,i}$ ,  $i \in W$ . For a given set of vaccine-strains  $W(\subseteq V)$ , the overall vaccine efficacy estimand is thus defined as

$$VE_{w|0} := 1 - \frac{\sum_{i \in W} q_{0,i}^T}{\sum_{i \in W} q_{0,i}^C} = \left\{ \frac{\sum_{i \in W} q_{0,i}^C \cdot VE_i}{\sum_{i \in W} q_{0,i}^C} \right\} \quad (1)$$

The rightmost equation expresses overall efficacy as a weighted average of strain-specific efficacies, the weights being the hazards of colonisation. In a strain-specific analysis, i.e., when  $W = \{i\}$ , it holds naturally that  $VE_{\{i\}|0} = VE_i$  as in the left part of (A3) and (B3).

The second estimand considers vaccine efficacy based on transitions from set  $\bar{V}_0$ , i.e., from state 0 or any of the singly-colonised states with a non-vaccine strain. This seems a natural extension of the “susceptibility set” from state 0 to states against which the vaccine does not have a biological effect mechanism. The set of target states  $\mathcal{W}$  now includes all states to which there is a direct transition (“arrow”) from any of the states in  $\bar{V}_0$ . To define a unique, time-independent hazard of transition between the two collections of states ( $\bar{V}_0$  and  $\mathcal{W}$ ), we need in addition to specify a mixing distribution for the time spent among the states in  $\bar{V}_0$ . It proves useful to apply the stationary distribution. For a given set  $W$  of target strains, the alternative overall vaccine efficacy estimand is then defined as:

$$VE_{w|\bar{V}_0} := 1 - \frac{\sum_{[h] \in \bar{V}_0} \bar{p}_{[h]|\bar{V}_0}^T \left( \sum_{[k] \in \mathcal{W}} q_{[h][k]}^T \right)}{\sum_{[h] \in \bar{V}_0} \bar{p}_{[h]|\bar{V}_0}^C \left( \sum_{[k] \in \mathcal{W}} q_{[h][k]}^C \right)}, \quad (2)$$

where  $\bar{p}_{[h]|\bar{V}_0}$  denotes the stationary probability of state  $[h]$ , conditioned upon being in  $\bar{V}_0$ .

Expression (2) is also a weighted average of strain-specific efficacies (see Web Appendix A). The two definitions thus fully agree for any strain-specific efficacy as in (A3). Moreover, for the overall efficacy they will yield similar numerical values if between-strain competition is homogeneously symmetric across all strains or acquisition of double colonisation is not common (see Web Appendix A).

Analogous to (1), we also define vaccine efficacy against the non-vaccine strains as the relative reduction in the overall hazard of acquisition of any of the non-vaccine strains from the uncolonised state:

$$VE_{\bar{V}_0} := 1 - \frac{\sum_{j \in \bar{V}} q_{0,j}^T}{\sum_{j \in \bar{V}} q_{0,j}^C} \quad (3)$$

By the assumption  $VE_j = 0$  for any non-vaccine strain, the theoretical value of  $VE_{\bar{V}_0}$  is 0. The estimand can naturally be restricted to any individual non-vaccine strain.

**Illustration cont'd**

In the four strain example, the overall vaccine efficacy against the two vaccine strains, based on definition (1), is  $VE_{\{1, 2\}|0} = (0.07 \cdot 0.7 + 0.06 \cdot 0.4) / (0.07 + 0.06) = 0.56$ . Based on definition (2), the overall vaccine efficacy  $VE_{\{1, 2\}|\{0, 3, 4\}}$  is also 0.56, calculated as

$$1 - \frac{\bar{p}_{0|\bar{v}_0}^T ((1-0.7)0.07 + (1-0.4)0.06) + \bar{p}_{\{3,4\}|\bar{v}_0}^T \cdot 0.4 ((1-0.7)0.07 + (1-0.4)0.06)}{\bar{p}_{0|\bar{v}_0}^C (0.07 + 0.06) + \bar{p}_{\{3,4\}|\bar{v}_0}^C \cdot 0.4(0.07 + 0.06)},$$

where  $(p_{0|\bar{v}_0}^T, p_{\{3,4\}|\bar{v}_0}^T) = (0.253, 0.134) / 0.387$  and  $(p_{0|\bar{v}_0}^C, p_{\{3,4\}|\bar{v}_0}^C) = (0.197, 0.103) / 0.3$  are the conditioned stationary probabilities of states 0 and {3, 4} in the vaccinees and controls. In this model, the competition parameter has the same value (0.4) for all pairs of strains and the overall efficacies based on (1) and (2) are thus equivalent (cf. Web Appendix A).

Throughout, we assume no interference among study subjects, and that the rate of exposure to colonisation does not change. This would be plausible if a small proportion of the population took part in the trial, such as a portion of infants. Under randomisation, expected exposure to colonisation would be the same in vaccinees and non-vaccinees. However, even under these circumstances, the hazards of acquiring non-vaccine strains would differ between vaccinees and non-vaccinees if not conditioned on current state of colonisation, because total time at-risk is distributed differently among the states of colonisation. The very motivation of this research is to elucidate the definition and estimation of vaccine efficacy in such situations.

**4. Cross-sectional estimation of vaccine efficacy against colonisation**

Here we present the main results of this paper by formulating expressions for  $VE_{acq}$  in (1) and (2) in terms of the stationary distribution, the stationarity requirement being imposed separately on the unvaccinated and vaccinated groups. Estimation of vaccine efficacy can then be based on these expressions, relying solely on cross-sectional measurement of colonisation. The following results are valid under model A. Derivation of results, proofs, and applicability under model B are shown in Web Appendix A. Let  $p_{[k]}^C$  and  $p_{[k]}^T$  denote the stationary probabilities of state  $[k] \in \mathcal{S}$  in the non-vaccinated (C) and vaccinated groups (T).

**Result 1**

The overall vaccine efficacy against a collection of vaccine strains  $W (\subseteq V)$ , as defined by (1), can be expressed in terms of the stationary probabilities as

$$VE_{w|0} = 1 - \frac{p_w^T / p_{\bar{v}_0}^T}{p_w^C / p_{\bar{v}_0}^C} \equiv 1 - \frac{\left( \sum_{[k] \in W} p_{[k]}^T \right) / \left( \sum_{[k] \in \bar{v}_0} p_{[k]}^T \right)}{\left( \sum_{[k] \in W} p_{[k]}^C \right) / \left( \sum_{[k] \in \bar{v}_0} p_{[k]}^C \right)}. \quad (4)$$

Expression (4) for  $VE_{w|0}$  has the form 'one minus an odds ratio'. The odds of being vaccinated (T vs. C) are compared between those colonised with the target strains and those colonised with the non-vaccine strains. The possible target states  $W$  include all singly-

colonised states for any target strain. The reference states  $\bar{v}_0$  involve all states in which no vaccine-strain is involved.

Instead of using  $\bar{v}_0$  as the reference set, it is possible to restrict the comparisons to state 0 (uncolonised). With this choice and using the non-vaccine strains as the target set, we obtain the following expression for vaccine efficacy against the non-vaccine strains to estimate (3):

$$VE_{\bar{v}_0} = 1 - \left\{ \left( \sum_{[k] \in \bar{V}} p_{[k]}^T \right) / p_0^T \right\} / \left\{ \left( \sum_{[k] \in \bar{V}} p_{[k]}^C \right) / p_0^C \right\}. \quad (5)$$

If the model is correctly specified, the theoretical value of (5) is 0, so the estimate of  $VE_{\bar{v}_0}$  should be close to zero (see Web Appendix A).

## Result 2

The overall vaccine efficacy against a collection of vaccine strains  $W (\subseteq V)$ , as defined by (2), can be expressed in terms of the stationary probabilities as

$$VE_{w|\bar{v}_0} = 1 - \frac{p_{\mathcal{W}}^T / p_{\bar{V}_0}^T}{p_{\mathcal{W}}^C / p_{\bar{V}_0}^C} \equiv 1 - \frac{\left( \sum_{[k] \in \mathcal{W}} p_{[k]}^T \right) / \left( \sum_{[k] \in \bar{V}_0} p_{[k]}^T \right)}{\left( \sum_{[k] \in \mathcal{W}} p_{[k]}^C \right) / \left( \sum_{[k] \in \bar{V}_0} p_{[k]}^C \right)}. \quad (6)$$

As with  $VE_{W|0}$  in (4), the reference set  $\bar{v}_0$  includes all those states in which no vaccine-strain is involved. Instead of the singly colonised states  $W$ , the target states  $\mathcal{W}$  now include all those states in which one of the target strains is involved either alone or together with a non-vaccine strain (Figure 2).

Expressions (4)–(6) provide estimators in terms of proportions of the subjects in different states of colonisation in the stationary phase. Thus, under stationary conditions both strain-specific efficacies and overall efficacy can be estimated without expensive and invasive longitudinal observation using standard methods for estimating odds ratios.

## Illustration, cont'd

Table 2 demonstrates how to calculate estimates based on  $VE_{W|\bar{v}_0}$  in (6) from observed strain distributions.

## 5. The aggregated model and sufficient conditions

To analyse the robustness of the proposed estimators to departures from the assumptions in Section 2, we formulate a sufficient set of conditions that still allows valid estimation of  $VE_{acq}$  from cross-sectional data. Consider the partition of the original set of colonisation states into three aggregated states (Figures 2A and 2B). The partition consists of the reference set  $\bar{v}_0$ , i.e., the colonisation states, including state 0, for which the vaccine has no direct effect on the hazard of entering these states, the target set  $\mathcal{W}$ , and the “rest”  $\mathcal{R}$ . We specify a new Markov model for the three aggregated states by defining the hazards of transition from one aggregate state to another in terms of the underlying transition hazards as follows:



$$q_{U,P} = \sum_{[h] \in U} \bar{p}_{[h]|U} \left( \sum_{[k] \in P} q_{[h][k]} \right), \quad U, P \in \{\bar{\mathcal{V}}_0, \mathcal{W}, \mathcal{R}\}, U \neq P, \quad (7)$$

where  $\bar{p}_{[h]|U}$  denotes the stationary probability of the state  $[h]$ , conditioned upon the aggregate state  $U$ , and  $q_{[h][k]}$  is the transition hazard from state  $[h]$  to state  $[k]$  (see Deng et al. (2009) for theoretical justification).

Under the aggregated model, it is natural to define another  $VE_{\text{acq}}$  for susceptibility of acquisition of strains in set  $W$  as

$$VE_{W|\bar{v}_0} = 1 - q_{\bar{\mathcal{V}}_0, \mathcal{W}}^T / q_{\bar{\mathcal{V}}_0, \mathcal{W}}^C, \quad (8)$$

i.e., the relative reduction in the hazard  $q_{\bar{v}_0, W}$  due to vaccination. This estimand is equivalent to (2) (see Web Appendix A for details). In addition, expression (6) for  $VE_{W|\bar{v}_0}$  can be interpreted as based on the aggregated model since the stationary distribution of the aggregated process is obtained by summing up the respective stationary probabilities under model  $A$ . It follows that, for any choice of the target strains, it is possible to formulate a smaller set of conditions directly in the context of the aggregated model that allows cross-sectional estimation of  $VE_{\text{acq}}$ . In the following we summarise these conditions.

For  $W = V$ , i.e., when estimating overall efficacy against all vaccine strains, the aggregated model has a simple chain structure of three states:  $\bar{v}_0 \leftrightarrow W \leftrightarrow \mathcal{R}$  (Figure 2A). It is then easy to show by the reversibility of chain models that the only condition required for cross-sectional estimation is the similarity of clearance hazards of the aggregate states  $W$  in the vaccinees and controls:

$$(B5) \quad q_{\mathcal{W}, \bar{\mathcal{V}}_0}^C = q_{\mathcal{W}, \bar{\mathcal{V}}_0}^T$$

Likewise, if the reverse hazards  $q_{W, \mathcal{R}}$  and  $q_{\mathcal{R}, W}$  satisfy the symmetry condition ( $q_{W, \mathcal{R}} / q_{\mathcal{R}, W} = q_{\bar{v}_0, \mathcal{R}} / q_{\mathcal{R}, \bar{v}_0}$ ), the essential model structure at stationarity is again a chain between three states:  $W \leftrightarrow \bar{v}_0 \leftrightarrow \mathcal{R}$  and (B5) is sufficient to estimate vaccine efficacy against *any* collection of vaccine strains ( $WC \ V$ ). Only when  $W$  is a strict subset of  $V$  and the competition is not symmetric, the aggregated model has the triangle structure (Figure 1B), and the conditions are more restrictive. Specifically, they are (B1)–(B4) for model  $B$  with three states corresponding to  $\bar{v}_0$ ,  $W$ , and  $\mathcal{R}$  (see Web Appendix A for more details).

## 6. Simulations, sensitivity analyses, and data analyses

### Simulated examples

We conducted a simulation study with four vaccine and five non-vaccine strains under model  $A$  to examine the performance of the proposed cross-sectional estimators. We chose the hazards of colonisation,  $q_{0, i}$ ,  $i = 1, \dots, 9$ , and the hazard of clearance  $\mu$  so the stationary strain distribution mimics that typical of serotypes of *S. pneumoniae*, with prevalence ranging from approximately 20% for the most common type to < 2% for rare types. Repetitions of samples in a cohort of 1000 vaccinees and 1000 unvaccinated controls were taken at the stationary phase of the process satisfying conditions (A1)–(A4).

Estimates of  $VE_{\text{acq}}$  based on  $VE_{W|\bar{v}_0}$  or  $VE_{W|0}$  were similar (Table 3), with both producing estimates close to the true values. Coverage probabilities are almost equal to the nominal confidence interval (90%). The use of  $VE_{W|\bar{v}_0}$  produces somewhat more precise estimates because a larger portion of the data is used. The slightly negative bias in the point estimates is due to bias in odds-ratio estimates from finite samples (see Web Appendix C Table C1).

Table 4 presents estimates of  $VE_{acq}$  in a similar scenario, except the observed data for any individual included at most one strain. This was achieved by randomly dropping one strain in any simulated sample with two simultaneous strains (approximately 20% of positive samples) and using estimator (4). As a general finding, this type of symmetry in missing data does not appear to affect the efficacy estimates. The estimates were not biased under heterogeneous clearance rates across individuals either (data not shown).

The last column in Table 4 presents the estimates of  $VE_{acq}$  proposed by Rinta-Kokko et al. (2009) (HRK), based on simple odds ratios obtained by splitting the observations into two complementary categories for each choice of the target strain(s). Such estimates are straightforward to calculate in a setting in which observations do not record double colonisation. However, this estimator does not take into account interactions between strains, thus systematically underestimates the strain-specific efficacies. In particular, estimates for the non-vaccine strains are grossly negative. Only the overall efficacy against all vaccine-strains agrees with the true value, which follows directly from the fact that expression (6) is equivalent to the simple odds ratio in this case.

### Sensitivity to model assumptions

To study sensitivity of the cross-sectional estimates to departures from model assumptions, we performed a simulation study for 9 strains, similarly as in Table 3. To estimate overall vaccine efficacy, the only assumption required is (B5). When vaccine enhances clearance of vaccine strains, violating (B5), the mean of the cross-sectional estimator (6) was close to a combined efficacy  $1 - (1 - VE_{W|v_0})(1 - VE_D)$  where  $VE_D$  is one minus the ratio of mean duration of colonisation in vaccinees to that in controls (Web Appendix C Table C2). The combined efficacy could not be estimated as well if vaccine decelerates clearance, which, however, is an implausible assumption. The estimate based on (6) can thus be interpreted more generally as a summary measure of the effect of being vaccinated on susceptibility to acquisition and on duration of colonisation. If individuals are infectious throughout colonisation, this quantity is related to the transmission potential of the strain in question (Rinta-Kokko et al. 2009; Becker and Starczak 1998). Similar results were obtained in extensive simulation studies for strain-specific efficacy against the transmission potential under symmetric competition between strains.

The theoretical results of Section 5 further mean the cross-sectional estimators for overall and strain-specific efficacy are also applicable with any difference in the clearance rates of the target strain and those in the class “rest”, at least under symmetric competition. We verified this by simulations in which the ratio of clearance rates for the target and the rest was as large (small) as 2 (0.5) (Web Appendix C Table C3). The same finding applies to the combined efficacy when the vaccine affects clearance (data not shown).

### Analysis of two pneumococcal vaccine trials

We used the cross-sectional estimators of vaccine efficacy to analyse data from the two randomised, controlled trials of the effect of pneumococcal conjugate vaccination on colonisation (Table 1). In both studies, double colonisation had not been recorded and, consequently, we calculated estimates based on (4) (for the vaccine types) and (5) (for the non-vaccine types). These can be interpreted as estimates of  $VE_{acq}$  in those individuals susceptible to acquisition. In both studies, vaccine efficacy against colonisation is estimated to be in the range 0.4–0.8 for most serotypes, although confidence intervals are wide. The overall efficacy against the non-vaccine types does not differ from 0 in either of the studies. The HRK efficacy estimates based on the simple odds ratios that do not adjust for within-host replacement in the last column are consistently smaller, demonstrating that the adjustment for time at-risk makes a difference.

## 7. Discussion

In this paper we consider estimation of  $VE_{acq}$ , vaccine efficacy against susceptibility to asymptomatic infection by a microorganism with multiple interacting strains, from cross-sectional data. We propose two estimands of  $VE_{acq}$  which differ in the way they condition on being at-risk to acquire the target strains. Specifically,  $VE_{w|0}$  is based on the relative reduction in the overall hazard of acquiring the target strains in an uncolonised individual whereas  $VE_{w|\bar{v}_0}$  is based on the corresponding hazard in all those susceptible to acquiring the target strains. The first has an individual level interpretation and would ideally be estimated from longitudinal data in the context of a (Markov) transition model. The second has a population level interpretation as the average efficacy under stationary conditions. Both estimands can be used for any set of strains and are coherent in the sense that the efficacy against all target strains is a weighted average of the respective strain-specific efficacies.

In practice, estimates of  $VE_{w|0}$  and  $VE_{w|\bar{v}_0}$  are similar when double colonisation is uncommon. Moreover, we showed assuming symmetry in competition, the two estimands are equivalent in any strain-specific analysis. For better statistical efficiency, estimates of  $VE_{w|\bar{v}_0}$  seem therefore preferable on most occasions. Under quite general conditions, the estimators yield the combined efficacy against susceptibility to acquisition and duration of colonisation. In particular, no assumption is needed about the clearance rates of the target strains and the rest. If the overall clearance rate of the target strains is not affected by vaccination, both estimators collapse to  $VE_{acq}$ . When competition between the target strain and the other vaccine-strains is not very asymmetric, the key assumption is stationarity.

Pneumococcal vaccine trials are most often conducted in infants, in whom a relatively stationary colonisation is achieved in unvaccinated individuals a few months after birth (e.g. Granat et al., 2007). In vaccinees, such data are more scarce, but it is plausible that a similar time period is sufficient to achieve stationarity (Nohynek et al., 2008). In our simulations, taking samples six months after vaccination was enough for successful cross-sectional estimation of vaccine efficacy. If using the proposed method in a specific epidemiologic setting, the appropriate sampling time should be examined by simulation.

To differentiate the effect of vaccination on acquisition and clearance rates requires longitudinal data. Vaccine-induced changes in the average clearance rate of vaccine strains through direct biological enhancement by vaccination on clearance is not supported by available data in the pneumococcal context (O'Brien et al., 1997; Dagan et al., 2002). Vaccination could also affect average clearance rates by changing the within-host mixture of vaccine-strains. This situation would require large differences in strain-specific efficacy for susceptibility and a wide distribution of clearance rates across different strains. Such differences are not evident between the pools of pneumococcal vaccine and non-vaccine serotypes although differences exist among individual serotypes (e.g. Lipsitch et al., 2012). Nevertheless, in these situations, we show the combined efficacy against susceptibility and duration is estimable from cross-sectional data if vaccination does not markedly lengthen duration of vaccine-strain colonisation.

The difference between estimates of  $VE_{w|\bar{v}_0}$  and the simple HRK estimates can be considerable for strain-specific efficacies (Tables 1 and 4). For pivotal trials with colonisation endpoints,  $VE_{w|\bar{v}_0}$  appears more justifiable as a measure of the biological effect of the vaccine on transmission potential or susceptibility to acquisition. This estimand is also more likely comparable across epidemiological settings with differing levels of competing exposure by non-vaccine strains. By contrast, the simpler HRK estimate with an anticipated negative efficacy for non-vaccine strains may be more informative about the

extent of replacement colonisation. Furthermore, a difference between the HRK estimate and the adjusted one for overall non-vaccine strains could give an indication of between-strain competition.

The *a priori* partition of strains into vaccine and non-vaccine strains is important in choosing the appropriate reference set of colonisation states ( $\mathcal{V}_0$ ). Some non-vaccine pneumococcal serotypes, however, may be affected by the vaccine through cross-immunity (Dagan et al., 2002). In this case, we showed the comparison set can be restricted to a smaller set of non-vaccine strains or even to the uncolonised state. The same argument can be applied in trials in which the control vaccine is another pneumococcal vaccine with a possibly different composition of vaccine-strains compared to the experimental vaccine.

A potential caveat in all approaches to estimating vaccine efficacy for multiple strains is the sensitivity of detection is less than perfect. In this study, the robustness of the suggested aggregation was shown with regard to complete insensitivity to detect double colonisation. However, the situation could be different with differential sensitivity across strains and under asymmetric competition between strains.

The proposed method of cross-sectional measurement of hazard ratios has a close connection to risk-set sampling in nested case-control studies. In particular, hazard ratios can be estimated by selecting controls from those at risk of the event of interest at the incident times of the cases (Wacholder et al., 1998). In the current setting with many states, the at-risk set corresponds to states from which there are arrows to the target states (Lubin, 1986). Specifically, although the setting in this paper is prospective, it can be cast as a case-control study. The “cases” are *prevalent*, i.e., subjects colonised with the target strains, while the “controls” are those in states from which a transition to the target state is possible or, more generally, states upon which the exposure (vaccine) does not have a direct effect. In contrast to nested case-control studies, cases themselves must not be included in the risk set.

Open questions for future research include incorporating asymmetric competition, heterogeneous response to vaccination, and allowing hazard rates of colonisation to change during the study. In the meantime, we have demonstrated under broadly general conditions, cross-sectional data can be used to estimate the protective effect of vaccination on colonisation.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

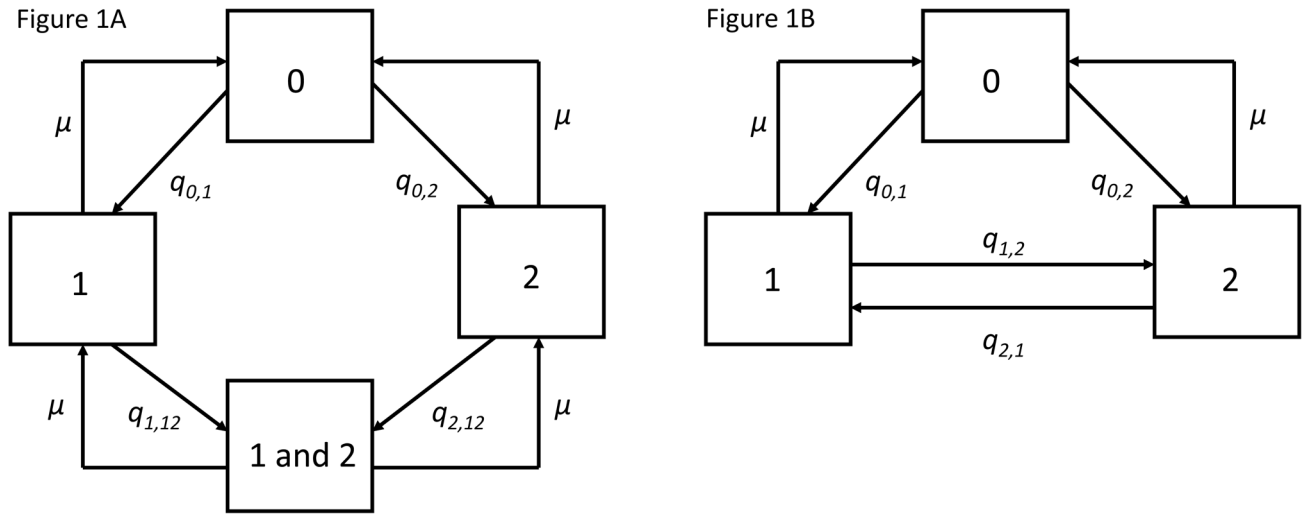
We thank Mikko Kosola, Sangita Kulathinal and Juha Mehtälä for useful discussions while preparing the manuscript. This study is part of the research of the PneumoCarr Consortium, funded by the Bill and Melinda Gates Foundation through the Grand Challenges in Global Health Initiative. MEH was supported by NIAID R37-AI032042.

## References

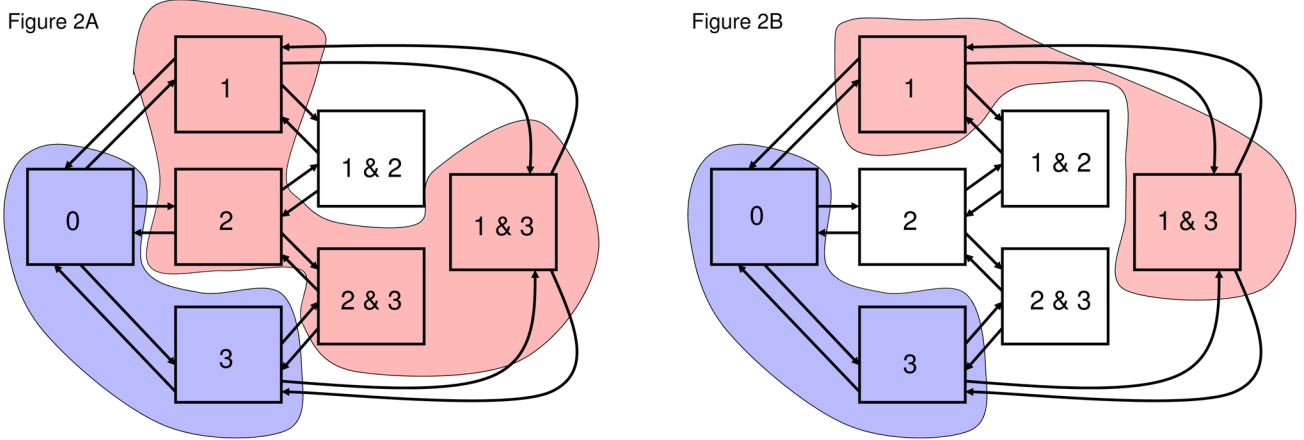
- Auranen K, Mehtälä J, Tanskanen A, Kaltoft M. Between-strain competition in acquisition and clearance of pneumococcal carriage - epidemiologic evidence from a longitudinal study of day-care children. *Am J Epidemiol.* 2010; 171(2):169–176. [PubMed: 19969530]
- Becker NG, Starczak DN. The effect of random vaccine response on the vaccination coverage required to prevent epidemics. *Math Biosci.* 1998; 154:117–135. [PubMed: 9949651]
- Dagan R, Givon-Lavi N, Zamir O, Sikuler-Cohen M, Guy L, Janco J, Yagupsky P, Fraser D. Reduction of nasopharyngeal carriage of *Streptococcus pneumoniae* after administration of a 9-

- valent pneumococcal conjugate vaccine to toddlers attending day care centers. *J Infect Dis.* 2002; 185:927–936. [PubMed: 11920317]
- Deng, K.; Sun, Y.; Mehta, PG.; Meyn, SP. An information-theoretic framework to aggregate a Markov chain. ACC'09 Proceedings of the 2009 conference on American Control Conference table of contents; IEEE Press Piscataway, NJ, USA. 2009. p. 731-736.
- Edwards MS. Group B streptococcal conjugate vaccine – a timely concept for which the time has come. *Human Vaccines.* 2008; 4(6):444–448. [PubMed: 18728401]
- Granat SM, Mia Z, Ollgren J, Herva E, Das M, Piirainen L, Auranen K, Mäkelä PH. Longitudinal study on pneumococcal carriage during the first year of life in Bangladesh. *Pediatr Infect Dis.* 2007; 26(4):319–324.
- Halloran, ME.; Longini, IM.; Struchiner, C. Design and Analysis of Vaccine Studies. Springer; 2010.
- Kaltoft MS, Skov Sørensen UB, Slotved HC, Konradsen HB. An easy method for detection of nasopharyngeal carriage of multiple *Streptococcus pneumoniae* serotypes. *J Microbiol Methods.* 2008; 75(3):540–544. [PubMed: 18801391]
- Kilpi, T.; Syrjänen, R.; Palmu, A.; Herva, E.; Eskola, J.; Mäkelä, PH. Parallel evaluation of 7-valent PncCRM on pneumococcal carriage and AOM. European Society for Paediatric Infectious Diseases Conference; 2001. <http://www.kenes.com/esp/2001/ESPIDAbstracts/200.htm>
- Lipsitch M. Vaccination against colonizing bacteria with multiple serotypes. *Proc Natl Acad Sci U S A.* 1997; 94(12):6571–6. [PubMed: 9177259]
- Lipsitch M, Dykes JK, Johnson SE, Ades EW, King J, Briles DE, Carlone GM. Competition among *Streptococcus pneumoniae* for intranasal colonization in a mouse model. *Vaccine.* 2000; 18(25):2895–2901. [PubMed: 10812233]
- Lipsitch M, Abdullahi O, D'Amour A, Xie W, Weinberger DM, Tchetgen Tchetgen E, Scott JA. Estimating rates of carriage acquisition and clearance and competitive ability for pneumococcal serotypes in Kenya with a Markov transition model. *Epidemiology.* 2012; 23(4):1–10. [PubMed: 22157298]
- Lubin JH. Extensions of analytic methods for nested and population-based incident case-control studies. *J Chron Dis.* 1986; 39(5):379–388. [PubMed: 3700579]
- Mbelle N, Huebner RE, Wasas AD, Kimura A, Chang I, Klugman KP. Immunogenicity and impact of nasopharyngeal carriage of a nonavalent pneumococcal conjugate vaccine. *J Infect Dis.* 1999; 180:1171–1176. [PubMed: 10479145]
- Melegaro A, Choi Y, Pebody R, Gay N. Pneumococcal carriage in United Kingdom families: estimating serotype-specific transmission parameters from longitudinal data. *Am J Epidemiol.* 2007; 166:228–235. [PubMed: 17517684]
- MMWR (CDC). Prevention of Pneumococcal Disease Among Infants and Children - Use of 13-Valent Pneumococcal Conjugate Vaccine and 23-Valent Pneumococcal Polysaccharide Vaccine Recommendations of the Advisory Committee on Immunization Practices (ACIP). *MMWR Morb Mortal Wkly Rep.* 2010 Dec 10.:59.
- Nohynek H, Mäkelä PH, Lucero M, Ugpo J, Auranen K, Forsyth S, Sombrero L, Herva E, Kaijalainen T, Nissinen A, Sanvictores D, Esparar G, Tallo V. ARIVAC Consortium. The impact of 11-valent pneumococcal conjugate vaccine (PCV) on nasopharyngeal carriage of *Streptococcus pneumoniae* (Pnc) in Philippine children. Abstracts of the 6th international symposium on pneumococci and pneumococcal diseases. 2008:abstract P3–134.
- O'Brien KL, Millar EV, Zell ER, Bronsdon M, Weatherholtz R, Reid R, Becenti J, Kvamme S, Whitney CG, Santosham M. Effect of pneumococcal conjugate vaccine on nasopharyngeal colonization among immunized and unimmunized children in a community-randomized trial. *J Infect Dis.* 2007; 196:1211–1220. [PubMed: 17955440]
- Tan LKK, Carlone GM, Borrow FRC. Advances in development of vaccines against *Neisseria meningitidis*. *New Engl J Med.* 2010; 362(16):1511–1520. [PubMed: 20410516]
- Rinta-Kokko H, Dagan R, Givon-Lavi N, Auranen K. Estimation of vaccine efficacy against acquisition of pneumococcal carriage. *Vaccine.* 2009; 27:3831–3837. [PubMed: 19490983]
- Struchiner CJ, Halloran ME, Brunet RC, Ribeiro JMC, Massad E. Malaria vaccines: lessons learnt from field trials. *Cadernos de Saúde Pública.* 1994; 10(Suppl 2):310–326. [PubMed: 15042221]

Wacholder S, Silverman DT, McLaughlin JK, Mandel JS. Selection of controls in case-control studies. I Principles. *Am J Epidemiol.* 1992; 135(9):1042–1050. [PubMed: 1595690]



**Figure 1.**  
**1A. Model A of colonisation.** The model has two strains and four states of colonisation. The states are uncolonised (0), colonised with strain 1, colonised with strain 2, and colonised with both strains (1 and 2). The model is governed by eight transition hazards, which fulfil conditions (A1) and (A2) if  $q_{1,12}/q_{2,12} = q_{0,2}/q_{0,1}$ . In the text, this model is considered for an arbitrary number of strains, separately for the vaccinated ( $T$ ) and control ( $C$ ) individuals.  
**1B. Model B of colonisation.** The model has two strains and three states of colonisation. The states are uncolonised (0), colonised with strain 1, and colonised with strain 2. The model is governed by six transition hazards, which fulfil conditions (B1) and (B2) if  $q_{1,2}/q_{2,1} = q_{0,2}/q_{0,1}$ . In the aggregated model, states 0, 1 and 2 correspond to aggregated states  $\mathcal{V}_0$ ,  $\mathcal{W}$  and  $\mathcal{R}$  (see Section 5).



**Figure 2.** Partition of states for estimation of vaccine efficacy. The figure presents the seven states of colonisation in model *A* for three strains. Strains 1 and 2 are vaccine strains and strain 3 is a non-vaccine strain, i.e.,  $S = \{1, 2, 3\}$  and  $V = \{1, 2\}$ . **2A. Overall vaccine efficacy.** The figure shows the partition for estimation of overall vaccine efficacy against strains 1 and 2. Here  $\mathcal{W} = \{1, 2, (1, 3), (2, 3)\}$  and  $\bar{\mathcal{V}}_0 = \{0, 3\}$ . The complement  $\bar{\mathcal{R}} = \{(1, 2)\}$  consists of the (only) doubly-colonised state with two vaccine strains. The aggregate model has a chain structure  $\bar{\mathcal{V}}_0 \leftrightarrow \mathcal{W} \leftrightarrow \bar{\mathcal{R}}$ . **2B. Strain-specific vaccine efficacy.** The figure shows the partition for vaccine efficacy against strain 1. Here  $\mathcal{W} = \{1, (1, 3)\}$ ,  $\bar{\mathcal{V}}_0 = \{0, 3\}$ , and  $\bar{\mathcal{R}} = \{2, (1, 2), (2, 3)\}$ . Generally, for an individual strain  $i \in S$ , define  $\mathcal{T}_i = \{i\} \cup \{(i, j); j \in \bar{V}, j \neq i\}$  as the set of those states in which strain  $i$  is involved either alone or with any of the non-vaccine strains. Then  $\mathcal{W} = \bigcup_{i \in V} \mathcal{T}_i$  and  $\bar{\mathcal{V}}_0 = (\bigcup_{i \in \bar{V}} \bar{\mathcal{T}}_i) \cup \{0\}$ .



**Table 1**

Vaccine efficacy against colonisation in two vaccine studies. **Upper part.** Data from a randomised, double-blind vaccine study on the effect of an experimental 9-valent pneumococcal conjugate vaccine on colonisation in Soweto, South Africa (Mbelle et al., 1999). Altogether 500 infants were recruited; 481 completed the 9-month follow-up. Children received either pneumococcal vaccine (Wyeth Lederle Vaccines, n=242) or placebo (n=239). The vaccine was given at 6, 10 and 14 weeks of age and the estimates of vaccine efficacy are based on measurements of colonisation at 9 months of age. Estimates are shown individually for seven of the nine vaccine serotypes (serotypes 1 and 5 were not found at all in the study), for all vaccine serotypes and for all non-vaccine serotypes. **Lower part:** Data from a randomised, double-blind vaccine study on the effect of a 7-valent pneumococcal conjugate vaccine on colonisation in Finland (Kilpi et al., 2001). Children received either of two pneumococcal vaccines (Wyeth Lederle Vaccines, n=831, or Merck Sharp and Dohme, n=835), or HBV vaccine (Merck Sharp and Dohme, n=831) as a control. The vaccine was given at 2, 4, 6 and 12 months of age and the estimates of vaccine efficacy are based on measurements of colonisation at 18 months of age. Estimates are shown individually for the seven vaccine serotypes (in the South African dataset, there were no samples with two of the vaccine serotypes, 1 and 5), for all vaccine serotypes, and for all non-vaccine serotypes. For both studies, the estimates of vaccine efficacy are determined against all vaccine serotypes and non-vaccine serotypes using estimators (4) and (5) for  $VE_{W|0}$  and  $VE_{\bar{V}|0}$ , respectively. For comparison, the last column shows estimates of vaccine efficacy (HRK), based on simple odds ratios (Rinta-Kokko et al., 2009; see text)

Serotype	Number of samples		Estimates of vaccine efficacy		
	Vaccinees	Controls	$VE_{W 0}$ (90% CI <sup>†</sup> )	HRK (90% CI <sup>†</sup> )	
<b>South Africa</b>					
4	3	5	0.54 (-0.54, 0.86)	0.41 (-0.97, 0.82)	
6B	8	28	0.78 (0.57, 0.89)	0.74 (0.49, 0.87)	
9V	1	2	0.62 (-1.89, 0.95)	0.51 (-2.71, 0.93)	
14	3	6	0.62 (-0.24, 0.88)	0.51 (-0.58, 0.85)	
18C	2	0	-	-	
19F	19	32	0.55 (0.25, 0.73)	0.45 (0.09, 0.67)	
23F	7	14	0.62 (0.17, 0.83)	0.52 (-0.04, 0.78)	
<b>Vaccine types</b>	43	87	0.62 (0.46, 0.74)	0.62 (0.46, 0.74)	
Non-vaccine types	87	58	-0.26 (-0.81, 0.12)*	-0.75 (-1.44, -0.26)	
0	112	94			
<b>Finland</b>					
4	1	2	0.77 (-0.76, 0.97)	0.75 (-0.87, 0.97)	
6B	44	39	0.47 (0.24, 0.64)	0.45 (0.20, 0.62)	
9V	4	4	0.53 (-0.50, 0.85)	0.50 (-0.60, 0.84)	
14	13	10	0.39 (-0.22, 0.70)	0.35 (-0.30, 0.68)	

Serotype	Number of samples		Estimates of vaccine efficacy		
	Vaccinees	Controls	VE <sub>W10</sub> (90% CI) <sup>†</sup>	HRK (90% CI) <sup>†</sup>	HRK (90% CI) <sup>†</sup>
<b>18C</b>	6	4	0.30 (-1.03, 0.76)	0.25 (-1.17, 0.74)	
<b>19F</b>	44	32	0.36 (0.05, 0.56)	0.32 (0.00, 0.54)	
<b>23F</b>	47	34	0.35 (0.06, 0.56)	0.32 (0.01, 0.53)	
<b>Vaccine types</b>	159	125	0.41 (0.27, 0.52)	0.41 (0.27, 0.52)	
Non-vaccine types	268	111	-0.16 (-0.42, 0.06) <sup>*</sup>	-0.25 (-0.52, -0.02)	
0	1176	564			

<sup>\*</sup> For vaccine efficacy against non-vaccine strains, the reference state is 0 and estimator (5) was used.

<sup>†</sup> CI = confidence interval.

**Table 2**

Estimation of vaccine efficacy from cross-sectional data. The table presents one simulated data set for 1000 vaccinated and 1000 control (non-vaccinated) individuals at the stationary phase in model A for two vaccine strains (1 and 2) and two non-vaccine strains (3 and 4), with hazards of colonisation 0.07, 0.06, 0.05, and 0.04 per time unit. The hazard of clearance ( $\mu$ ) was 0.17 per time unit, and the relative hazards for acquisition of double carriage ( $q_{ij}$ ,  $i, j$ ) was 0.4 for all  $i, j$ ; see eq. (A1). For each strain, the numbers of observations are given for the singly and the three doubly-colonised states with that strain. The last column shows how to calculate vaccine efficacy against individual and all vaccine-strains (VS) by using estimator (6) for  $VE_{W|V_0}$  and vaccine efficacy against individual and all non-vaccine strains (NVS) by using estimator (5). The true values of vaccine efficacy are given in the first column. The true value of the overall efficacy against the vaccine-strains is  $VE_{V|V_0} = VE\{1;2\} = 0.56$ , based on definition (2) and equation (E1)

True efficacy	States of colonisation	Numbers of samples in vaccinees	Numbers of samples in controls	Estimates of vaccine efficacy $VE_{W V_0}$
0.70 ( $VE_1$ )	1+1&2+1&3+1&4	61 + 4 + 7 + 6	162 + 29 + 18 + 9	$1 - \frac{(61+7+6)/(522+260+13)}{(162+18+9)/(403+207+9)} = 0.70$
0.40 ( $VE_2$ )	2+1&2+2&3+2&4	107 + 4 + 12 + 8	136 + 29 + 17 + 10	$1 - \frac{(107+12+8)/(522+260+13)}{(136+17+10)/(403+207+9)} = 0.39$
0.00 ( $VE_3$ )	3+1&3+2&3+3&4	131 + 7 + 12 + 13	111 + 18 + 17 + 9	$1 - \frac{131/522}{111/403} = 0.09^*$
0.00 ( $VE_4$ )	4+1&4+2&4+3&4	129 + 6 + 8 + 13	96 + 9 + 10 + 9	$1 - \frac{129/522}{96/403} = -0.04^*$
0.56 (all VS)	VS+VS& NVS+VS&VS <sup>†</sup>	168 + 33 + 4	298 + 54 + 29	$1 - \frac{(168+33)/(522+260+13)}{(298+54)/(403+207+9)} = 0.56^{††}$
0.00 (all NVS)	NVS+VS & NVS+NVS& NVS <sup>‡</sup>	260 + 33 + 13	207 + 54 + 9	$1 - \frac{260/522}{207/403} = 0.03^*$
0	-	522	403	-

\* For efficacy against non-vaccine strains, the reference state is 0 and estimator (5) was used.

<sup>†</sup> VS&NVS is the number of subjects colonised with a vaccine-strain and a non-vaccine strain, VS&VS is the number of subjects colonised with two vaccine-strains, NVS&NVS is the number of subjects colonised with two non-vaccine strains.

<sup>††</sup> For the vaccine-strains, the overall estimate is based on eq. (6) for  $VE_{V|V_0} = VE\{1, 2\} = 0.56$ . An alternative estimate given by eq. (4) is  $VE_{V|V_0} = VE\{1, 2\} = 1 - (168/(522+260+13))/(298/(403+207+9)) = 0.56$ .

**Table 3**

Performance of cross-sectional estimators of vaccine efficacy assuming perfect detection of the singly and doubly-colonised states. The results are based on 1000 simulated data sets, each with 1000 vaccinees and 1000 controls. The model consisted of 4 vaccine (shown in boldface) and 5 non-vaccine strains, with hazards of colonisation 0.0075, 0.003, 0.002, 0.001, 0.005, 0.003, 0.0009, 0.0005, and 0.0003 per time unit (day). The relative hazard for acquisition of double carriage ( $q_{i, ij/q_{0, i}}$ ) was 0.4 for all  $i, j$ . The hazard of clearance  $\mu$  was 0.017 per day. The simulation was started from state 0 (uncolonised) for each individual and the sample for calculating the vaccine efficacy was taken at day 183. The true vaccine efficacy was 0.70 against strains 1 and 3, and 0.40 against strains 2 and 4. The true value of the overall efficacy is  $VE_{V|V_0} = 0.61$ , based on definition (2) and equation (E1). The program code to analyse an example data with 9 strains is provided in a Web Appendix B.

Strain	True efficacy	No. of singly and doubly colonised vaccinees <sup>†</sup> (% of all samples)	Controls	Estimates of vaccine efficacy (SE log(OR); 90 % cov. prob.) <sup>††</sup>
		Vaccinees		$VE_{V V_0}$
<b>1</b>	<b>0.7</b>	66 + 19(8.5)	166 + 61(22.7)	0.70(0.14; 0.91)
<b>2</b>	<b>0.4</b>	53 + 16(6.9)	66 + 31(9.7)	0.39(0.18; 0.90)
<b>3</b>	<b>0.7</b>	18 + 6(2.4)	44 + 22(6.6)	0.70(0.27; 0.90)
<b>4</b>	<b>0.4</b>	17 + 6(2.3)	22 + 11(3.3)	0.38(0.30; 0.91)
5	0	146 + 34(18.0)	110 + 46(15.6)	-0.01(0.14; 0.90)*
6	0	88 + 24(11.2)	66 + 31(9.7)	-0.01(0.18; 0.92)*
7	0	26 + 9(3.5)	20 + 10(3.0)	-0.06(0.31; 0.92)*
8	0	15 + 5(2.0)	11 + 6(1.7)	-0.13(0.42; 0.90)*
9	0	9 + 3(1.2)	6 + 4(1.0)	-0.18(0.56; 0.92)*
<b>Vaccine strains</b>	<b>0.61</b>	154 + 41(19.5)	298 + 96(39.4)	0.61(0.11; 0.89)
Non-vaccine strains	0	284 + 55(33.9)	213 + 82(29.5)	0.00(0.11; 0.91)*
0		501(50.1)	378(37.8)	-

\* For efficacy against non-vaccine strains, the reference state is 0 and estimator (5) was used.

<sup>†</sup>The mean number of samples in the 1000 simulated data sets.

<sup>††</sup>The average standard error of the log odds ratio, based on the 1000 simulated data sets. The coverage probability of a 90% confidence interval, based on the 1000 simulated data sets.

Performance of cross-sectional estimation of vaccine efficacy under imperfect detection of double colonisation. The simulated complete data were the same as used in Table 3. Consequently, all doubly-colonised states were recorded as singly-colonised, with 50% chance of either of the colonising strains being detected.

**Table 4**

Strain	True efficacy (VE <sub>t</sub> )	No. of colonised samples <sup>†</sup> (% of all samples)	Controls	VE <sub>HT</sub> <sup>‡</sup>	Estimates of vaccine efficacy (SE log(OR); 90% cov. prob.) <sup>††</sup>	HRK
		Vaccinees				
1	0.7	75(7.5)	196(19.6)	0.70(0.15; 0.90)	0.66(0.14; 0.78)	
2	0.4	61(6.1)	82(8.2)	0.42(0.18; 0.90)	0.27(0.18; 0.72)	
3	0.7	21(2.1)	55(5.5)	0.71(0.27; 0.89)	0.64(0.27; 0.84)	
4	0.4	20(2.0)	28(2.8)	0.40(0.30; 0.90)	0.24(0.30; 0.83)	
5	0	163(16.3)	133(13.3)	0.07(0.14; 0.85) <sup>*</sup>	-0.28(0.13; 0.40)	
6	0	100(10.0)	81(8.1)	0.07(0.16; 0.86) <sup>*</sup>	-0.27(0.16; 0.59)	
7	0	31(3.1)	25(2.5)	0.05(0.28; 0.88) <sup>*</sup>	-0.27(0.28; 0.82)	
8	0	17(1.7)	14(1.4)	0.00(0.38; 0.91) <sup>*</sup>	-0.34(0.38; 0.86)	
9	0	10(1.0)	8(0.8)	-0.06(0.50; 0.91) <sup>*</sup>	-0.42(0.49; 0.88)	
<b>Vaccine strains</b>	<b>0.61</b>	177(17.7)	361(36.1)	0.62(0.11; 0.88)	0.62(0.11; 0.88)	
Non-vaccine strains	0	321(32.1)	261(26.1)	0.08(0.11; 0.80) <sup>*</sup>	-0.33(0.10; 0.11)	
0		502(50.2)	378(37.8)			

<sup>\*</sup> For efficacy against non-vaccine strains, the reference state is 0 and estimator (5) was used.

<sup>†</sup> The mean number of samples in the 1000 simulated data sets.

<sup>††</sup> The average standard error for the log odds ratio, based on the 1000 simulated data sets. The coverage probability of a 90% confidence interval, based on the 1000 simulated data sets.