# 10

# Household-based studies

## 10.1 Concepts of household studies

Household studies are important for studying the effects of vaccines, transmission, and the natural history of infection. In Chapter 2 we introduced vaccine efficacy parameters that require conditioning on exposure to infection. Household studies were used as the basis for defining exposure to infection in vaccine studies as early as the 1930's in evaluating the efficacy of pertussis vaccines (Kendrick and Eldering 1939). Historically the interest focused on evaluating the protective effects of vaccination. The relative risk of developing illness in vaccinated compared to unvaccinated susceptibles exposed to cases in their household was the basis of estimating the protective effects, $\mathrm{VE}_{S,p}$, or $\mathrm{VE}_{SP,p}$. In observational studies, evaluating vaccine efficacy under conditions of household exposure can help reduce bias generated by unequal exposure in vaccinated and unvaccinated people. In recent years, the vaccine effect on the ability to transmit the infection in vaccinated infected people compared to unvaccinated infected people, $\mathrm{VE}_I$, has gained attention. An additional measure of interest is the overall reduction in transmission if both the infective and the susceptible are vaccinated compared to if neither are vaccinated, $\mathrm{VE}_T$. Considering the estimates of VE based on the relative secondary attack rates, there are three main unstratified vaccine effects:

$$VE_{S.1/.0} = 1 - \frac{SAR_{.1}}{SAR_{.0}} \ , \quad VE_{I1./0.} = 1 - \frac{SAR_{1.}}{SAR_{0.}} \ ,$$
$$VE_T = 1 - \frac{SAR_{11}}{SAR_{00}} \ . \tag{10.1}$$

If one stratifies on the vaccine status of the infective person or the susceptible person, then there are four further stratified measures of $\mathrm{VE}_S$ and $\mathrm{VE}_I$:

$$VE_{S01/00} = 1 - \frac{SAR_{01}}{SAR_{00}}, \quad VE_{S11/10} = 1 - \frac{SAR_{11}}{SAR_{10}},$$
$$VE_{I10/00} = 1 - \frac{SAR_{10}}{SAR_{00}}, \quad VE_{I11/01} = 1 - \frac{SAR_{11}}{SAR_{01}}. \tag{10.2}$$

Equations (10.1) and (10.2) give the three main unstratified and three stratified vaccine effects conditonal on exposure to infection data. Despite being widespread, household studies of vaccine effects have not generally been used for primary licensure efficacy trials. Household studies are sometimes nested within randomized, controlled studies and provide secondary analyses.

The analysis is generally based on the relative transmission probability, $p$, or relative secondary attack rate (SAR), between the vaccinated and unvaccinated individuals of interest. The SAR is a special case of the transmission probability. The secondary attack rate is the expected proportion of susceptibles who become infected when exposed to an infectious person. In the secondary attack rate, the contact between the infectious susceptible persons may be defined as occuring over some time period, such as the duration of infectiousness or over the period of the study. For example, the *household SAR* is the probability that a susceptible individual living in the same household with an infectious person during his or her period of infectiousness will become infected (Fine, et al, 1988; Orenstein, et al, 1988). The SAR is a proportion, not a rate. The index case in a household is the case that draws attention to the household and leads to ascertainment of the household. The index case is also often the first, or primary, case in the household, but not necessarily. A case that occurs too soon after the primary case to have resulted from infection by the primary case is called a co-primary case.

Households are the most common form of transmission unit used. It allows easy identification of contacts between a case and susceptibles, and families are convenient units of study. Many other settings are also used as transmission units in studies and analyses that condition on exposure to infection. These include sexual partnerships, classrooms, schools, school buses, day care centers, and workplaces, among others. Here we talk mostly about household studies, but many of the studies and analyses are applicable with possibly slight modification to other transmission units as well. We use the term "household" for general small transmission units. The term household is much easier for exposition than is "transmission unit".

Often but not always the household exposure studies are nested in a study that has the primary analysis based on one of the unconditional measures of vaccine efficacy, such as $\mathrm{VE}_{S,IR}$ or $\mathrm{VE}_{S,CI}$. In these studies, when an exposure is determined to have occurred, for instance, when a sibling of a vaccine study participant has a case of pertussis, then the outcomes are evaluated in a secondary analysis.

In addition to evaluating vaccine efficacy, household studies have been used to learn about the transmission and natural history of many infections. Aspects of the natural history studied in households include the transmissibility, the incubation and latent periods, the duration of infectiousness, and the serial interval between cases (Hope-Simpson 1952; Bailey 1957). Household studies have also been used to evaluate other interventions, such as post-exposure prophylaxis with influenza antiviral agents (Welliver et al 2001; Hayden et al. 2004). Exposure to an infectious case within a household can be used as

a natural challenge study, for example when studying immunological correlates of protection. In many analyses of household studies, the households are assumed to be independent of one another, so that susceptible contacts are exposed only by the first case within the household. When the statistical model assumes that the households are embedded within a community, the analysis allows estimation of the risk of being infected in the community as well as the risk of infection by exposure to a case within the household and the vaccine effects at both levels.

In this and the following two chapters, we consider households studies not only for evaluating vaccine effects, but in a broader context. Some of these concepts may be useful for future vaccine studies. The household- and school-based pneumococcal carriage studies were conducted as a prelude to introduction of the pneumococcal vaccines. This chapter provides several examples of household studies and discusses general design considerations. Design considerations include how the households are ascertained, whether the cases are ascertained on infection status or symptomatic cases, and whether the studies are randomized or observational. The data structure and follow-up period can depend on whether the infection results in immunity that lasts at least as long as the study period, such as in influenza, colds, or measles, or whether a person can experience repeated episodes of infection, carriage or disease during the study, such as pneumococcal nasopharyngeal carriage. Chapters 11 and 12 cover methods of analysis in more detail. Chapter 11 presents several methods for analyzing data assuming that households are embedded in communities. Chapter 12 presents methods of analysis assuming that households are independent.

## 10.2 Pertussis Vaccination

### 10.2.1 History

Household exposure studies have long been used to evaluate pertussis vaccination. Pertussis vaccines were developed in the 1920's and the first hopeful results were observed in the Faroe Islands in the early 1920's (Madsen 1933; Medical Research Council 1951). Most pertussis vaccines were based on killed whole cells until the 1980's. Concern about efficacy and adverse effects of whole-cell pertussis vaccines resulted in some countries to discontinue recommending its use. For example, Sweden completely discontinued pertussis vaccination in 1979 because the efficacy seemed to be negligible (Trollfors 1981). A new generation of acellular vaccines was developed as an alternative to the killed whole-cell ones. In the 1980's and early 1990's, considerable interest in evaluating the relative efficacy of the two types generated a number of papers on how methodological compared to biological effects of vaccines affected the efficacy estimates. Fine and Clarkson (1987) and Fine et al (1988) give a thoughtful review of sources of variability in pertussis vaccine efficacy

estimates. They compare estimates based on controlled trials, cohort studies, case-control studies, and secondary attack rate studies. Efficacy estimates were often lower in household studies, possibly due to more intense and prolonged exposure.

In countries that did not recommend pertussis vaccination, trials of the efficacy of the new vaccines could be conducted with a placebo arm. In countries that recommended use of the whole cell pertussis vaccines, it was unethical to have a placebo arm, and the two vaccines had to be compared head to head. Children not in the study who were not vaccinated could be followed and provide an unvaccinated study arm as part of an observational study. Pertussis vaccine is generally combined with the diptheria and tetanus toxoids and given three to four times early in the first year of life. The vaccine without the pertussis component is denoted DT, and with it is denoted DTP. We present several examples of pertussis vaccine studies in households.

### 10.2.2 Michigan, USA

Kendrick and Eldering (1939) report on a study of pertussis immunization in children between 8 months and <5 years (<6 years for a short time at the beginning) in Grand Rapids, Michigan, USA, and surrounding areas from March 1, 1934 to November 1, 1937. Although the study was not randomized, efforts were made to create a control group comparable to the test group. Children receiving the vaccine were self-selecting. They obtained the vaccine by presenting themselves at the city immunization clinics. As children were immunized, comparable children were selected at random from a population-based roster to match the vaccinated children on age and district. House visits were made by nurses to all participants initially at 3 to 4 month intervals, but after November 1935 at 2 month intervals. Public health and other sources of reports of whooping cough cases were followed up as well.

The diagnoses in the study were primarily based on detailed clinical histories. Kendrick and Eldering discuss the difficulties associated with diagnosing an attack of pertussis with certainty, particularly one in which the usually accepted clinical criteria are lacking or at least not prominent. The difficulty of choosing the best case definition for pertussis persists even today.

The main analysis was based on the relative number of cases per person-years at risk in the vaccinated group compared with the control group (Figure 1.2 and equation (2.4)). However "from the beginning, one important objective in the study was to obtain as exact information as possible with regard to exposures to pertussis and subsequent related attacks" (Kendrick and Eldering 1939, page 146). They had clearly established definitions of exposures. To be considered an exposure, the source case had to have a written case history with diagnosis made on the same basis as the study participants. The contact had to be recorded. Different levels of exposure were defined. The levels of exposure were (1) definite in their own household, (2) definite in other households, (3) indefinite, and (4) no exposure history. To be considered definite,

an exposure had to occur within 21 days of onset of the source case. A maximum incubation period of 30 days was assumed. Definite exposures in other households had to be of at least 30 minutes duration. Indefinite exposures could occur under less intimate conditions, such as outdoors or after the 21st day, but no later than the 35th day of onset of the source case. The data are shown in Figure 1.3 and the vaccine efficacy estimate based on definite household exposure is in equation (2.5).

### 10.2.3 Niakhar, Senegal

Active population surveillance has been conducted since 1983 in Niakhar, Senegal, a sub-Saharan rural community of 30 villages. The community is very homogeneous, composed of Sereer peasant families, living in compounds, the residential unit for extended families. As part of many research components (Garenne and Cantrelle 1998), pertussis was under prospective and active surveillance (Préziosi, et al. 2002). As a result, for each child information was available not only on pertussis illnesses and vaccination but also on contacts. Extended families were under longitudinal observation beginning in March 1983, based on annual visits, and from 1987 to 1996, based on weekly visits to each compound. In addition, during pertussis vaccine trials 1990-1996 comparing whole cell to acellular vaccine, physicians collected biological samples from consenting suspected cases in the entire population, defined as having a cough lasting 8 days or more. The pertussis vaccine studies were conducted in the 1990's in accordance with the Helsinki Declaration (Préziosi et al 1997). The children who did not receive vaccination in the trials were under active surveillance as well. Samples included nasopharyngeal aspirates for isolating the bacteria and to detect DNA using PCR. Acute and convalescent blood samples were drawn to measure IgG titers to pertussis toxin (PT) or filamentous hemagglutinin (FHA) by ELISA. Surveillance for pertussis focused on children under age 15 years. All suspected cases and their co-residents were followed actively by a physician. The usual demographic data, including age, gender, hut, compound, hamlet and village were known for each child in the area. Pertussis vaccination status and dates of vaccination were also known. The primary analysis of the efficacy trials was based on unconditional vaccine efficacy parameters (Simondon, et al. 1997).

For each suspected case, the date of symptom onset, duration of cough, type of cough, a wide range of symptoms, results of each biologic diagnostic test done, and physician diagnosis were recorded. Focusing on the year 1993, an epidemic year that produced a large number of cases and extensive exposure to pertussis, Preźiosi and Halloran (2003) and Halloran, et al (2003) analyzed the data to estimate not only $\mathrm{VE}_S$ but also $\mathrm{VE}_I$ and $\mathrm{VE}_T$ for pertussis. Preźiosi and Halloran (2003) considered a number of different case definitions and the relation to estimated $\mathrm{VE}_S$, $\mathrm{VE}_I$ and $\mathrm{VE}_T$. Halloran et al (2003) considered different statistical methods for the secondary attack rate analysis (see Chapter 12.2) using just one case definition. In the latter

paper, a case of pertussis was defined as requiring clinically, at least 21 days of cough with paroxysms and biologicially, either *B. pertussis* isolated from a nasopharyngeal aspirate or significant increase or decrease in PT or FHA antibodies as measured by ELISA or presence of a bacteriologically confirmed case in the same compound within 28 days. The latter criterion is called an epilink.

Preźiosi and Halloran (2003) chose the compound as the transmission unit within which it was assumed that susceptibles were exposed to infection by the first case in the unit. The compound is the "home", i.e., the residential unit where individuals make privileged contacts and where random mixing is a reasonable assumption. Indeed, for these reasons, the compound is the transmission unit of choice in African rural settings such as here (Garenne, et al 1993; Aaby et al 1996).

A potentially infectious contact, or exposure, was defined as a susceptible living in the same compound during the infectious period of the index case. Exposed susceptibles were children with no history of pertussis living in a compound with an index case. Onset of pertussis symptoms was assumed to be the onset of infectiousness, thus the latent period equals the incubation period. Co-primaries were those cases whose onset of cough was $<7$ days after that of the index case, assumed to be too soon after the index case to have been infected by the index case. To allow for uncertainty in duration of infectiousness, a secondary case was defined as a case whose date of onset was $\geq 7$ days after that of the index case and less than a variable cutoff, specifically none, 56, 42, or 28 days.

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

A total of 518 of the 1800 compounds (29%) were detected as having potential cases of pertussis in 1993. In 189 (36%) of those compounds, pertussis was confirmed. They represented 232 primary and co-primary cases and 1217 susceptibles. Among those were excluded compounds with co-primary cases (n=33 [17%]), compounds with no susceptibles (n=5[3%]), and compounds with a partially vaccinated primary case (n=42[22%]). Thus a total of 109/189 (58%) of the qualifying compounds was eligible for analysis. The 109 compounds represented 109 primary cases and 790 susceptibles, of whom 152 ([19%] were partially vaccinated and 638 [81%] were either unvaccinated or completely vaccinated. Table 10.1 gives the data and SAR's using different cutoffs. The result of at least one biological confirmation criterion was available in over 97% of the suspected cases meeting our clinical definition. From the same study, Preźiosi and Halloran (2003b) estimated the effect of pertussis vaccination on clinical severity $\mathrm{VE}_P$ (Chapter 9).

**Table 10.1.** Number of exposed susceptibles, secondary cases, and secondary attack rates (SAR) by vaccination status of the index case and the exposed susceptible children and cutoff for counting secondary cases (from Halloran et al 2003).

| | Exposed susceptibles and secondary cases | | | | | |
|---|---|---|---|---|---|---|
| | Vaccinated | | Unvaccinated | | Combined | |
| Index case | cases/exposed | SAR | cases/exposed | SAR | cases/exposed | SAR |
| Vaccinated | | | | | | |
| cutoff: none | 11/127 | 0.09 | 9/67 | 0.13 | 20/194 | 0.10 |
| 56 days | 10/127 | 0.08 | 6/67 | 0.09 | 16/194 | 0.08 |
| 42 days | 10/127 | 0.08 | 5/67 | 0.07 | 15/194 | 0.08 |
| 28 days | 3/127 | 0.02 | 3/67 | 0.04 | 6/194 | 0.03 |
| Unvaccinated | | | | | | |
| cutoff: none | 61/246 | 0.25 | 73/198 | 0.37 | 134/444 | 0.30 |
| 56 days | 55/246 | 0.22 | 67/198 | 0.34 | 122/444 | 0.27 |
| 42 days | 52/246 | 0.21 | 66/198 | 0.33 | 118/444 | 0.27 |
| 28 days | 41/246 | 0.17 | 52/198 | 0.26 | 93/444 | 0.21 |
| Combined | | | | | | |
| cutoff: none | 72/373 | 0.19 | 82/265 | 0.31 | 154/638 | 0.24 |
| 56 days | 65/373 | 0.17 | 73/265 | 0.28 | 138/638 | 0.22 |
| 42 days | 62/373 | 0.17 | 71/265 | 0.27 | 133/638 | 0.21 |
| 28 days | 44/373 | 0.11 | 55/265 | 0.21 | 99/638 | 0.16 |

### 10.2.4 England

During World War II, several investigations were undertaken by the Whooping-cough Immunization Committee of the Medical Research Council to assess the prophylactic value of pertussis vaccination, with disappointing results. Between 1946 and 1950, the committee conducted an essentially randomized, controlled trial in children between 6 and 18 month when recruited. They tested five batches of vaccine from three manufacturers, two from the Michigan Department of Health, two from Glaxo Laboratories, and one from Parke Davis and Co. in ten separate field trials (Medical Research Council 1951). Each child in the study was visited monthly by a nurse-investigator. Information was obtained on exposure to pertussis, incidence of upper-respiratory track disease, other immunizations, and other childhood diseases. If it was found by the visit or routine report by the parent that a child had been exposed to pertussis or had developed suspicious symptoms, repeated visits were made and the mother was asked to take notes as well.

A total of 6,710 children completed the trial, with 3,358 in the vaccinated and 3,352 in the unvaccinated group. In the vaccine group, there 149 cases in 102,961 child-months at risk, and in the unvaccinated group, there were 687 cases in 102,180 child-months at risk, a risk ratio of 1 to 4.6. The results give a $\mathrm{VE}_{S,IR} = 1 - 1.45/6.72 = 0.78$, [95% CI 0.74,0.82]. Analysis of information on the exposures of children to pertussis was divided into two categories. First,

**Table 10.2.** Total number of cases of pertussis and secondary attack rates by type of exposure according to vaccine group from the study by the Medical Research Council in England 1946–1950.

| | Home exposure | | | Other exposure | | | No |
|---|---|---|---|---|---|---|---|
| Vaccination status | No of exposures | No of cases | Rate (%) | No of exposures | No of cases | Rate (%) | exposure history |
| Vaccinated | 203 | 37 | 18.2 | 566 | 47 | 8.3 | 65 |
| Unvaccinated | 173 | 151 | 87.3 | 561 | 213 | 38.0 | 323 |

home exposures were children exposed in their own home to one or more siblings, and second, other exposures were children exposed in "day nurseries, in nursery schools, at parties, in cinemas, in buses, and while playing outside the home with other children." In this study, the number of exposures was recorded, not the number of children exposed, as some children were exposed more than once. Table 10.2 gives the summary data, not broken down by the 10 areas and 5 vaccine batches. When analyzed by vaccine batch, the two vaccines from the Michigan Department of Health gave a considerably greater degree of protection than the other three.

After this study, England continued to monitor efficacy of pertussis vaccine. As the controversy over the vaccine continued, a fresh assessment was made. During an outbreak that began in 1977, from January 1978 through June 1980, England undertook a national assessment of the efficacy of pertussis vaccination in 21 area health authorities (PHLS Epidemiologic Research Laboratory 1982). The 21 areas comprise about one-quarter of the total health authorities in England at that time. Case notification rates for children with three doses of DTP or three doses of DT were studied in that period. The vaccination status both of the population under 6 years of age and of the notified cases was provided from computer records by each area health authority (AHA). Home visits by nurses and health visitors from the AHA were made to notified cases to assess the severity of the case, the family circumstances, and to take perinasal swabs. Information was collected on age, sex, history of pertussis in the distant past, and history of recent illness that could have been pertussis. Particular attention was given to children under 6 years of age. A subsequent home visit about six weeks later was also made to record symptoms in contacts under six years. Nurses were asked to report all cases of cough whether or not associated with typical paroxysms. A household contact who developed spasmodic cough was considered a case. The original analysis included only two-child households. About 90% of the notified cases were visited.

In the DTP group, a total of 2261 cases were notified in about 250,163 child-years at risk (0.9%). In the DT group, a total of 9,515 cases were notified in 187,595 child years at risk (5.1%) over the course of the study. Efficacy,

**Table 10.3.** Secondary attack rates in home contacts according to age and vaccine group (from PHLS Epidemiological Research Laboratory 1982)

| Age of contact (years) | 3 DTP | | | 3 DT | | | Relative rate DTP:DT |
|---|---|---|---|---|---|---|---|
| | Contacts | No of cases | Rate (%) | Contacts | No of cases | Rate (%) | |
| $0- < 1$ | 28 | 12 | 43 | 56 | 34 | 61 | 1:1.4 |
| $1- < 2$ | 108 | 35 | 32 | 399 | 316 | 79 | 1:2.5 |
| $2- < 3$ | 97 | 36 | 37 | 384 | 299 | 78 | 1:2.1 |
| $3- < 4$ | 108 | 34 | 31 | 284 | 170 | 60 | 1:1.9 |
| $4- < 6$ | 476 | 92 | 19 | 428 | 165 | 39 | 1:2.0 |

$VE_{S,IR}$, based on the total number of cases for each year of birth was greater than 0.80. However, the analysis based on the secondary attack rates in households was lower in the study. Table 10.3 shows the relative secondary attack rates in two-child families in which symptoms in the contact began at least one week after those of the index case. Efficacy was consistently around 0.50, except in the children less than one year, where the number of cases is small. In this study, the co-primaries were those within seven days of the index case and secondary cases were those that occurred within about 42 days of the index case and at least seven days after the index case. The efficacy was higher with a more severe case definition, reaching 71 percent in children with 10 paroxysms or more.

Fine, et al (1988) reanalyzed this study and considered why estimates of pertussis vaccine efficacy might be lower in household contact studies than when assessed in cohort analyses in general populations. They restricted their analyses to households with at least one child under 6 years of age. The primary case was defined as the first recent case in the household, which in many households was not the index case. Co-primaries were defined as cases within one week of the primary case. Incidence cases were those that occurred more than one week after the primary cases. These included more than potentially secondary cases. Incidence cases were further divided into retrospective, prospective, and current incidence cases depending on whether they occurred before, after, or around the time of the initial visit to the household. The analysis included 9,242 households with 10,406 contacts, of whom 6,436 (61.8%) developed pertussis at the same time or after symptom onset in the primary case. The 1,520 co-primary cases were excluded from further analysis. A surprising 94% of all incidence cases were retrospectively ascertained.

There were two key findings. First, vaccine efficacy was lower, though not significantly, in retrospectively than in prospectively ascertained cases. The overall, age standardized efficacy was 0.35 (95% CI 0.25–0.44) in retrospectively ascertained cases, and 0.59 (95% CI 0.42–0.70) in prospectively ascertained cases. Secondly, the efficacy was lower, though not significantly,

in contacts exposed to vaccinated primary cases than in contacts exposed to unvaccinated primary cases. This latter finding is not consistent with the biological argument that the bacterial exposure from a vaccinated case would be lower than from an unvaccinated case (Préziosi and Halloran 2003a). They speculate that it could be due to household clustering of vaccine failures or false positive diagnoses.

### 10.2.5 Sweden

After cessation of pertussis vaccination in 1979 in Sweden, pertussis became endemic again (Romanus et al 1987). Thus it was possible to conduct randomized, placebo-controlled trials of pertussis vaccination in Sweden. A trial of two acellular pertussis vaccines compared with placebo was conducted in Sweden 1986–1987. The efficacies were lower than expected, which could have been due to more sensitive case ascertainment, so further efficacy trials were planned directly comparing the acellular with whole cell vaccines. Several pertussis vaccine trials were conducted in Sweden in the 1990's.

In a double blind, placebo-controlled trial in the Göteborg area of western Sweden, 3450 infants were randomized to vaccination with DT or the same DT with pertussis toxoid at 3, 5, and 12 months of age. The study children were born between June 1991 and May 1992 (Trollfors et al 1995). Trollfors et al (1998) were interested in estimating the indirect protection of close contacts of the children in the vaccine trial. A household study was nested within the primary efficacy study described in Chapter 6.4.2. Parents and siblings in households were followed for a median of two years starting 30 days after the third vaccination up to January 31, 1995. The numbers of older siblings of the DTP and DT were 938 and 965, of younger siblings 514 and 523, and of parents 3237 and 3229, respectively. The vaccination status of parents and siblings of the study children was not recorded. This is an example of the mini-community design (Section 10.7.5).

Later acellular pertussis vaccine candidates contained further antigens. Storsaeter et al (1998) did a study to evaluate immunological surrogates of protection after household exposure to pertussis. The idea was to use household exposure as a natural challenge experiment in studying surrogates of protection. The study was nested in a primary efficacy study (Gustafsson et al 1996). The household study is reported in Chapter 15.4.3. Further examples of studies of the efficacy of acellular pertussis vaccination after household exposure are Trollfors et al (1997) in Sweden and Schmitt et al (1996) in Germany.

## 10.3 Influenza

Longitudinal household studies have a long history in the study of transmission of influenza and other acute respiratory diseases. Household studies of

influenza have generally not been used for estimating vaccine efficacy, though they have been used for evaluating the effects of post-exposure prophylaxis of influenza antiviral agents. We present a number of household-based studies of influenza transmission for their historical significance and to promote future household-based studies of influenza and other respiratory diseases. We present the household-based studies of influenza antivirals to illustrate further methodological issues.

### 10.3.1 Seattle USA

Intensive surveillance of Seattle, Washington, USA families with school-age children for influenza virus infections was conducted from 1975 to 1979 (Fox et al 1982a). The study followed the Virus Watch method that basically involves continuing virologic surveillance of families. The Viral Watch in Seattle began by recruiting families with newborn infants in 1965 to 1969 with a focus on respiratory and enteric viruses detectable by cell culture methods and that were not well understood at that time. The Virus Watch method was specifically adapted for the study of influenza viruses to yield a better description of their behavior. Families with at least one child were recruited in Fall 1975 (Group I) or Fall 1976 (Group II) and followed for three years. In Group I, 112 families were recruited, and in Group II, 116 familes were recruited. By the 1978–1979 season, the families had dwindled to 44 and 73, yielding a total of 639 family-seasons of observation over four influenza virus epidemic seasons.

The protocol required collection of blood samples by venipuncture at four-month intervals, information concerning onset and manifestation of symptoms, and duration of illness in any family member, using illness records kept by the mother. Nose-throat swab specimens for virus isolations were to be collected from all family members on a regular basis, bi-weekly or, during influenza outbreaks, weekly, particularly when onset of a new case occurred. The plan was quite ambitious and could not be fully implemented. Many illnesses were missed, though there is no way to estimate how many. Between 9% (Group I) and 13% (Group II) of reported illnesses had no specimens collected, while between 26% (Group I) and 32% (Group II) of illnesses were recognized only because specimens were collected. Fox et al (1982b) analyzed the pattern of infection in invaded households and the relation of age and prior antibody to occurrence of infection and related illness. Susceptibility to each type or subtype was rigorously defined so that the resulting secondary attack rates would reflect virus infectivity. Susceptibles were defined on the basis of a pre-episode hemagglutination-inhibiting antibody titer of $1{:}\leq 20$ for A/H3N2 virus and $1{:}\leq 10$ for A/H1N1 and type B viruses. Of 102 contacts susceptible to A/H3N2, 53% became infected when exposed in the household. Of 147 contacts susceptible to A/H1N1, 44% were infected when exposed. Of 55 contacts susceptible to type B, 47% became infected.

**Table 10.4.** Observed distribution of influenza A(H3N2) infections in 1977-1978 and 1980-1981 combined epidemics in Tecumseh, Michigan (from Addy et al 1991).

| No. | No. of susceptibles per household | | | | |
|---|---|---|---|---|---|
| infected | 1 | 2 | 3 | 4 | 5 |
| 0 | 110 | 149 | 72 | 60 | 13 |
| 1 | 23 | 27 | 23 | 20 | 9 |
| 2 | | 13 | 6 | 16 | 5 |
| 3 | | | 7 | 8 | 2 |
| 4 | | | | 2 | 1 |
| 5 | | | | | 1 |
| Total | 133 | 189 | 108 | 106 | 31 |

### 10.3.2 Tecumseh, USA

Active community surveillance of acute respiratory illness took place in Tecumseh, Michigan, during the five year period 1976–1981 (Monto, et al 1985). Beginning in October 1976, recruitment over a three month period resulted in 1,000 individuals, approximately 10% of the community being under surveillance by the end of December. The households were recruited in a stratified manner until the required number was reached. Initially there were no restrictions on elibility. Because of attrition, further recruitment was necessary. In 1978 the requirement that a family have at least one child of school age or younger was added. Then in 1979, families were recruited at the birth of the child until the end of the study in 1981. Throughout the five years of the study, families on surveillance were called weekly to identify the onset of acute illness. Specimens for virus isolation were collected when an illness was reported within two days of symptom onset. Blood specimens were collected from all on surveillance at six-month intervals. In addition, specimens for virus isolation were collected by Tecumseh physicians from patients with febrile respiratory illness. Table 10.4 contains a summary of the distribution of influenza A(H3N2) infections in 1977-1978 and 1980-1981 combined epidemics in Tecumseh, Michigan given in Addy et al (1991). Addy et al (1991) give the household frequency data in Table 10.4 stratified by age group 0–17 years and 18+ years as well. Table 10.5 contains a summary of the data stratified by age group and pre-season antibody titer (Longini et al 1988). The criterion for classifying individuals as susceptible is a preseason hemagglutination inhibition test detecting no antibody in a dilution of 1 in 128 or less. People with higher titers were considered immune and are not included in the tables. Households with more than five susceptibles were deleted from all analyses. Longini et al (1988) give the household level frequency data stratified by pre-season antibody level and age group group.

**Table 10.5.** Infection attack rates by pre-season antibody titer level stratified by age group: influenza A(H3N2) epidemic seasons 1977-1978 and 1980-1981 combined in Tecumseh, Michigan (from Longini et al 1988).

| Pre-season antibody titer | Infection status | | | |
|---|---|---|---|---|
| $(1:x)$ | No. infected | No. not infected | Total | Attack rate |
| Children (0–17 years) | | | | |
| Low level ($x < 8$) | 100 | 200 | 300 | 0.333 |
| High level ($8 \leq x \leq 64$) | 20 | 180 | 200 | 0.100 |
| Total | 120 | 380 | 500 | 0.240 |
| Adults (18+ years) | | | | |
| Low level ($x < 8$) | 96 | 440 | 536 | 0.179 |
| High level ($8 \leq x \leq 64$) | 42 | 402 | 444 | 0.095 |
| Total | 138 | 842 | 980 | 0.141 |

### 10.3.3 Cleveland, USA

A large longitudinal 10 year study of illness of families in Cleveland, Ohio, USA was conducted from January 1, 1948 through May 31, 1957 (Dingle at al 1964). The were two primary objectives of the study. The first was to answer questions such as how much illness actually, occurs, what is the etiology of the illnesses, how important is the family unit in spreading the illness, do families have a characteristic pattern of illness, and do individuals and families vary in susceptibility to illness. The second objective was to study specific diseases, using clinical, epidemiological, and laboratory results. The study had four parts. First, illnesses or events occurring in each individual and family were observed and recorded. Second, known entities such as streptococcal infections, influenza, or noninfectious diseases were differentiated and their behavior studied. Third, possible entities of unknown etiology were investigated. Fourth, problems such as the spread of infectious agents in the population, evaluation of therapeutic or prophylactic agents, and the occurrence of noninfectious processes were studied. Stable, middle class families with at least one child were recruited. Extensive medical examinations were done on each family when it entered the study and at regular intervals, either 6-month or one-year in children, and annually in adults. Records were kept by each mother, who notified the investigators at the time of each illness, however minor. Each family was visited weekly by a field worker, who obtained a throat culture from each member of the household. The family physician was called when necessary. During the study, an epidemic of poliomyelitis occurred in 1952, and stool specimens were collected. Some diseases, such as chickenpox, were recognized more reliably than others.

A total of 96 families and 443 individuals were in the study at one time or another. In May 1957, the first reports of the new antigenic variant of

**Table 10.6.** Influenza attack rates by age as measured by virus isolation (Jordan et al 1958).

| Age groups | | Respiratory illness | | | | |
|---|---|---|---|---|---|---|
| | | | Test for virus | | Virus isolated | |
| (years) | No. | No. | No. | Per cent | No. | Per cent |
| 0–4 | 28 | 44 | 35 | 79.6 | 12 | 42.9 |
| 5–9 | 76 | 113 | 80 | 70.8 | 44 | 57.9 |
| 10–14 | 68 | 108 | 83 | 76.6 | 40 | 58.8 |
| 15+ | 17 | 27 | 19 | 70.4 | 8 | 47.1 |
| Adults | 119 | 100 | 71 | 71.0 | 22 | 18.5 |
| Totals | 308 | 392 | 288 | 73.5 | 126 | 40.9 |

influenza virus A occurred in Asia. In anticipation of the influenza pandemic, the Cleveland study was reactivated in September 1957. Sixty of the families agreed to participate again for collection of detailed clinical and epidemiologic data (Jordan et al 1958). Table 10.6 contains the influenza illness attack rates by age as measured by virus isolation during the Asian influenza pandemic in the 60 families.

### 10.3.4 Influenza Epigrippe, France

The Epigrippe study was conducted during the 1999-2000 influenza season in France (Carrat et al 2002). Households were recruited for follow-up by 161 general practitioners. In total 946 households were recruited. For a household to be included, a member of the household had to visit a general practitioner with a history of fever ($\geq 38^o$C) in the last 48 hours and respiratory signs. The household had to have at least one other member, everyone had to give consent to participate in the study, and the patient seeking care had to be the first case in the household and not be hospitalized as a result of the illness. In all index cases, nasal swabs were obtained at the first visit. Biological confirmation of influenza virus was by immunofluorescence test and/or culture and/or PCR. Households followed up with diaries of symptoms for 15 days after recruitment of the index case. Influenza was defined clinically in contacts. Of the 946 index cases, 510 tested positive for influenza virus. Follow-up information was obtained on 334 (65%) of the households with positive index cases. Cauchemez, et al (2004) analyzed the data that included the 334 confirmed index cases and households and 350 clinical influenza cases in 790 contacts. Influenza in symptomatic contacts was not confirmed biologically, nor was there any biological confirmation of possible asymptomatic infections. A case of influenza in the contacts was defined as having clinical influenza for at least one day.

**Table 10.7.** Some characteristics of the four studies as reported in the four papers.

| | Zanamivir | | Oseltamivir | |
|---|---|---|---|---|
| | Zan I<br>Hayden et al<br>2000 | Zan II<br>Monto et al<br>2002 | Osel I<br>Hayden et al<br>2004 | Osel II<br>Welliver et al<br>2001 |
| Centers | 15 | 59 | multi | 76 |
| Where | US, Canada,<br>UK, Finland | S. Africa, Europe<br>New Zealand,<br>NA, Australia | North<br>America,<br>Europe | North<br>America,<br>Europe |
| Study<br>period | Oct 98 – Apr. 99 | June 2000 – Apr. 2001 | 2000-01 season | 1998–99 season |
| Predominant<br>types | B (∼30% )<br>A(H3N2) | B (∼33%)<br>A(H1N1) (north)<br>A(H3N2) (south) | B (∼33%)<br>A(H1N1) | B (∼47%)<br>A(H3N2) |
| Randomized: | | | | |
| No. families (IC) | 337 (321) | 487 | 277 | 374 |
| No. contacts | 837 | 1291 | 812 | 962 |
| Inf. index cases*: | | | | |
| Control arm | | | | |
| Households (IC)† | 87 (81) | 153 | 84 | 79 |
| No. contacts | 215 | 398 | 228 | 206 |
| Treatment arm | | | | |
| Households (IC) | 78 (76) | 129 | 89 | 84 |
| No. contacts | 195 | 368 | 248 | 209 |

* includes only households with laboratory-confirmed index cases
† IC = index case

### 10.3.5 Influenza antivirals

Four randomized household-based studies of the efficacy of post-exposure prophylaxis in preventing clinical influenza in household contacts were conducted, two of zanamivir (Hayden et al. 2000; Monto et al. 2002), called Zan I and Zan II, and two of oseltamivir (Hayden et al 2004; Welliver et al 2001), called Osel I and Osel II (Halloran et al 2007). Table 10.7 contains a summary of some characteristics of the four studies. All four studies were household-based, multi-center, randomized, controlled trials, where treatment was randomized by household (cluster randomized design). Households with a suspected case of influenza illness were enrolled as a whole in each study. Assignment of the index case to treatment or control varied across the studies, resulting in differences in the effect measures estimated in each study. Ages for eligibility of index cases and contacts also varied across studies.

- Zan I (Hayden et al. 2000): Randomized, double-blind, placebo-controlled trial. Households were randomized to study drug (zanamivir) or placebo.

Index cases and eligible contacts within a household all received either drug or placebo. Children under age 5 did not receive study drug.

- Zan II (Monto et al. 2002): Randomized, double-blind, placebo-controlled trial. Households were randomized for eligible contacts to receive study drug (zanamivir) or placebo. Index cases did not receive antiviral therapy. Children under age 5 did not receive study drug.
- Osel I (Hayden et al. 2004): Randomized, open-label, trial. Households were randomized for eligible contacts to receive either antiviral post-exposure prophylaxis or antiviral treatment when illness developed (expectant treatment). All index cases received study drug (oseltamivir) treatment for 5 days. Children under 1 year were excluded from participating.
- Osel II (Welliver et al. 2001): Randomized, double-blind, placebo-controlled trial. Households were randomized for eligible contacts to receive study drug (oseltamivir) or placebo. Index cases did not receive antiviral therapy. Children under 12 years were excluded from participating as contacts, but could be (untreated) index cases.

In all four studies, the primary endpoint in the household contacts was laboratory-confirmed clinical influenza illness. A secondary endpoint was laboratory-confirmed influenza infection, whether symptomatic or asymptomatic. All four studies did extensive laboratory testing of the enrolled index cases and their contacts. Because contacts were tested for influenza infection regardless of whether they had symptoms, it is possible to estimate pathogenicity from the data (Chapter 9). Contacts were supposed to complete diary cards once or twice daily for 14 days or more, depending on the study, with details of symptoms and temperature. The definitions of clinical symptomatic influenza cases essentially included fever and symptoms, though they varied across the four studies. The period for inclusion of secondary cases in the original analyses varied across the studies.

Analogous to the vaccine efficacies in equations (10.2), from the appropriate $\mathrm{SAR}_{jk}$'s, in principle, we can estimate the stratified antiviral efficacies, $\mathrm{AVE}_S$, $\mathrm{AVE}_I$, and $\mathrm{AVE}_T$. Three main design issues are illustrated by these studies that are applicable for vaccine studies as well. First, household randomization restricts the efficacy parameters that can be estimated, discussed later in this chapter. Secondly, asymptomatic infections in contacts were ascertained, so that pathogenicity and the effect of prophylaxis on pathogenicity, $\mathrm{AVE}_P$, could be estimated. Third, each of the efficacies $\mathrm{AVE}_S$, $\mathrm{AVE}_I$, and $\mathrm{AVE}_T$ could be based on laboratory-confirmed influenza illness, $\mathrm{AVE}_{\cdot d}$, or simply laboratory-confirmed infection, $\mathrm{AVE}_{\cdot i}$, in the eligible contacts.

## 10.4 Measles vaccination

Measles vaccines are generally much greater than 90 percent efficacious against clinical disease. One of the considerations is at what age infants or children

should be vaccinated. Maternal antibodies transferred before birth protect very young infants and interfere with the live vaccine virus being able to induce an immune response in the infant. If vaccinated too young, vaccination will not be effective. On the other hand, if vaccinated too late, then the maternal antibody protection will have waned, and the child could easily contract measles before being vaccinated. In the US, vaccination against measles occurs between 12 and 15 months. However, in developing countries, this is often too late because exposure is more wide spread. Considerable research has been directed at understanding the optimal age to vaccinate infants in developing countries. In addition, new vaccines with high titers of vaccine virus were tried that were thought could induce antibodies at a younger age.

### 10.4.1 Niakhar, Senegal

The clinical efficacy of three measles vaccines was studied in a randomized trial in Niakhar, Senegal, in the same population described in Section 10.2.3. Garenne et al (1993) evaluated the efficacy of measles vaccines after controlling for the level of exposure to infection within the compounds. They conducted two analyses of efficacy, one based on the unconditional cases per person-time at risk, the other based on the secondary attack rate within compound. The first analysis was based on a randomized vaccine trial conducted from August 1987 to July 1990 to compare two high-titer vaccines, the Edmonston-Zagreb and the Schwarz, and the standard Schwarz. (Garenne et al 1991). The randomized trial covered the cohorts of children born between February 1987 and January 1989. The children were randomized into the three vaccine groups, with the two high-titer vaccines being administered at 5 months and the standard Schwarz at 10 months. The unvaccinated group were those children who were not available to be vaccinated on their scheduled day. An unvaccinated control arm was unethical. A total of 1,566 children were vaccinated, with vaccine coverage of 81.6% of the resident target population. The analysis controlling for the level of exposure within compound was nested in the randomized study.

Three measles outbreaks occurred during the study period. In the first, 27 cases occurred between May and September 1988, 161 cases between October 1988 and July 1989, and 413 cases between August 1989 and July 1990. When a family suspected a case of measles or a case was seen in the clinic, a specifically trained physician went to the compound. The physician visited the compound twice a week until the last case was cured. For serologic confirmation, an initial blood sample was obtained by fingerprick in susceptible children in the family during the first visit, with a second sample obtained from clinical cases at least four weeks after the onset of rash.

Exposure was defined as being susceptible (those who had never had measles) and being present in a compound where there was a clinical case of measles. Secondary cases were defined as those occurring in the same compound 7 to 18 days after the index case. The mean time lag between index

**Table 10.8.** Incidence and secondary attack rates of measles in a randomized trial of three measles vaccines in 30 villages 1987–1989, Niakhar, Senegal (Garenne et al 1993).

| | Prospective study | | | Compound exposure study | | |
|---|---|---|---|---|---|---|
| Group | Resident January 1, 1990 | Cases reported/ confirmed | Incidence rate per 1,000 p-yrs | Contacts | Cases reported/ confirmed | SAR (%) |
| Schwarz | 740 | 1/0 | 0.80 | 54 | 1/0 | 1.85 |
| HT EZ | 552 | 5/3 | 4.12 | 53 | 3/2 | 5.66 |
| HT Schwarz | 274 | 5/2 | 6.67 | 24 | 2/1 | 8.33 |
| Unvaccinated | 348 | 54/21 | 40.63 | 46 | 30/13 | 65.22 |

and secondary cases was 12.2 days, similar to that found in previous analyses (Hope-Simpson 1952; Bailey 1957). Different levels of exposure within compounds were defined using a linear score: 1 = living in a different compound; 2 = living in same compound but eating from a different kitchen; 3 = eating from the same kitchen but sleeping in a different hut; 4 = sleeping in the same hut. Reported clinical cases could be either directly or indirectly confirmed. Direct confirmation required fulfilling the clinical case definition and having at least a fourfold rise in HIA to measles virus during the acute phase. Indirect confirmation was by epilink, that is, when it occurred in a compound where another case was directly confirmed.

## 10.5 Pneumococcal carriage

Pneumococcal diseases are a major health problem all over the world. The etiologic agent is *Streptococcus pneumoniae* (Pnc), a bacterium surrounded by a polysaccharide (sugar) capsule. There are about 90 different serotypes of Pnc differentiated by the composition of the capsule. Pneumococcal bacteria are prevalent in populations. Generally the pneumococcal bacteria colonize the nasopharyngeal area without causing symptoms. Sympomatic disease can be either invasive or noninvasive. Invasive disease includes pneumonia, meningitis, and bacteremia with fever. Noninvasive disease includes otitis media and bronchitis. Generally, the cases of disease, especially invasive disease, are not considered infectious for others, at least not important for transmission. In contrast, the asympomatic carriers are considered to be the main sources of infection. People have the ability to acquire colonization in the nasopharynx and to clear it repeatedly without developing complete immunity. Given the numerous serotypes, a person may acquire one type of infection, clear it, then acquire either the same type or another.

The original pneumococcal vaccines were based on the polysaccharide capsule with up to 23 of the serotypes. The first was licensed in the US in 1977,

with an improved version in 1983. Immunogenicity was not great, so a new generation of conjugate vaccines was developed based on purified polysaccharide joined to a harmless variety of diphteria toxin. The conjugate pneumococcal vaccine was licensed in the US in 2000. These vaccines contain 7 to 11 serotypes and induce a T-cell dependent immune response. They have been shown to be effective in children and a strong population effect is being observed. In preparation for introducting the new vaccines, a series of household-based carriage studies were done in a number of different countries. The studies were to study the acquisition and clearance of the different serotypes, their relative prevalence, and possible difference in their acquisition and clearance rates. One question of scientific interest was whether vaccination against the vaccine serotypes would increase not only the relative but also absolute prevalence of nonvaccine serotypes.

In pneumococcal carriage studies, the time of onset and the time of clearance of carriage are not observed, so households are not generally ascertained on an index case. Households may be ascertained on some aspect of the index person, such as having a young infant in the household. Household members are examined at regular intervals to determine whether they are carrying the bacteria. Follow-up is active. The data are longitudinal, also called panel data, with repeated sampling of the same individuals at fixed, or nearly fixed, time intervals.

### 10.5.1 Finland

Auranen et al (2000) analyze data from the FinOM cohort study concerning the epidemiology of acute otitis media with a special emphasis on *Streptococcus pneumoniae* (Pnc) bacteria (Syrjänen et al 2001). Healthy unselected babies born to Finnish-speaking mothers and not previously immunized with a pneumococcal vaccine were consecutively enrolled at their first routine visit to a local well-baby clinic in Tampere, Finland between April 1994 and August 1995. Nearly all babies in Finland attend such clinics. During the enrollment period, 53% of the families with a newborn chose to participate in the study. The infants were followed for nasopharyngeal carriage of Pnc over a period of two years. Auranen et al (2000) analyzed a subset of 97 infants and their families for which carriage information was collected from all family members. The 97 infants were enrolled consecutively between December 1994 and May 1995.

During the follow-up, 14 younger siblings of the index children were born. All family members ($N = 370 + 14$) were examined for Pnc carriage when the index child was 2, 3, 4, 5, 6, 9, 12, 15, 18, and 24 months old, for a total of 10 time points over the two-year follow-up period. Time is defined for each family from birth of the index child. At each observation, the absence or presence of Pnc was identified for the seven Pnc serotypes that were to be included in the new vaccine. The proportion of recorded observations was 86% of the potential number, which is high for such extensive follow-up. In 40 of the 97

families, there was no observed carriage in anyone in the family during the follow-up period.

### 10.5.2 France

A five month longitudinal study of 3- to 6-year old children in 81 schools was conducted in France from January to May 2000 (Guillemot et al 2005). Children were examined for Pnc carriage using oropharyngeal swabs approximately once a month over a five month period (Figure 10.1). Oropharyngeal swabs are less sensitive than using nasopharyngeal swabs. The mean time between consecutive swabs was 37 days (sd 15 days). During the observation period 9,857 swabs were collected for serotyping. The 4,488 3- to 6-year old childen attending the schools represented 88% of the children in the area under study. Of these, 2,445 (55%) gave at least one swab. The mean number of swabs was four (range: one to five) among children providing at least one swab. All children attending the schools were included in the analysis as a density factor, even if they had not provided a single observation of follow-up (Cauchemez et al 2006). The analysis was restricted to the 16 serotypes isolated in at least 30 swabs in the selected schools. The analysis divided the serotypes into two groups, those contained in the seven-valent vaccine and those not. The study preceded the introduction of the vaccine into France, so all participating children were unvaccinated. Cauchemez et al (2006) analyzed this study using methods similar to Auranen et al (2000).

### 10.5.3 United Kingdom

A study of 121 preschool children $< 3$ years old and all household members was conducted in the United Kingdom during the follow-up period from October 2001 to July 2002 (Hussain et al 2005). Enrollment was through primary health care registers in Hertfordshire. Families were visited once a month over a 10 month period. All family members were examined for carriage using nasopharyngeal swabs. At least one swab was obtained from 489 individuals in 121 families for a total of 3,753 swabs, of which 932 (25%) were positive for Pnc. Melegaro et al (2004) modeled the household transmission similarly to Auranen et al (1996). However, they used maximum likelihood estimation to estimate the transition rates between carriage and noncarriage (Section 11.4.1).

### 10.5.4 Bangladesh

A study in a community-based project in a transitional area in Savar, Bangladesh enrolled 99 children born between May 2000 and April 2001 and their families (98 because 2 newborns were twins). (Granat et al 2007). The families were visited every two weeks until the index child was 4 months old,

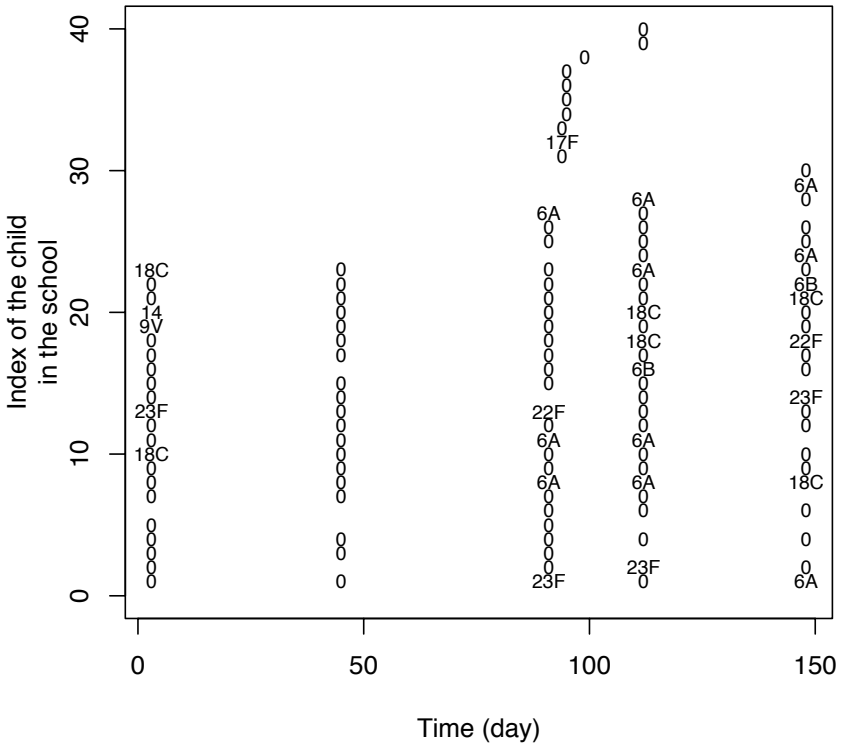


**Fig. 10.1.** Longitudinal data in a school participating in a pneumococcal carriage study in France (Guillemot et al 2005). A "0" represents a sample in which no pneumococcal serotypes was detected. The other symbols represent the pneumococcal serotype abbreviation of the detected bateria. (from Cauchemez et al 2006d)

then monthly up to 1 year of age, for a total of 16 visits. The goal of the study was to describe the development of pneumococcal carriage in a developing country setting. Swabs were taken from the infant and from other children and family members present and consenting during the visit. A total of 1,459 samples (92% of those planned) were collected from the 99 index children and 2,865 from the other family members. Approximately 50% of the infants had acquired pneumococcal carriage by 8 weeks of age. The point prevalence of pneumococcal carriage in the first 5 years was about 50% and declined thereafter to between 7 and 8% in adults.

## 10.6 Design Considerations

### 10.6.1 Transmission units and contacts

The scientific question of interest will influence the design of the study in households or other transmission units. The concept of a contact is very broad and must be defined in each particular study. The transmission mode of an infectious agent determines what types of contact are potentially infectious. Contacts can be defined between two individuals, or an individual and a vector. Contacts can be defined within small transmission units, such as households. Within small transmission units, mixing is often assumed to be random. A small transmission unit can also be defined as two individuals, such as a steady sexual partnership or a household with just two susceptible people. The definition of a contact within a study can depend on the definition of the transmission units. The individuals in a small transmission unit exposed to an infectious case can be thought of as a *minicohort* (Orenstein et al 1988) that has its own reference date for exposure to infection. An advantage is that vaccination status is less likely to change over the time of follow-up. A small transmission unit can also be thought of as a *minicommunity* if the indirect effects of vaccination of a fraction of the people in the transmission unit are of interest.

Different definitions of a potentially infective contact and transmission unit, for the same infectious agent, even within the same study, are possible. In a study of chickenpox transmission, a potentially infective contact could be defined as being in the same school on one day with someone with chickenpox. Alternatively, it could be defined as living in the same house during the presumed infectious period of the person with chickenpox. In the first case, the transmission unit is the school, and in the latter, it is the household. In the first case, the contact is defined over one day, and in the latter, it is defined over the entire infectious period. In tuberculosis, a contact could be defined as riding on the same bus with someone with open tuberculosis, or as living in the same household with someone with tuberculosis. In the former case, the transmission unit is the bus, and in the latter, it is the household.

There could be different definitions of a contact for one definition of transmission unit. In an HIV study, a potentially infective contact could be defined as each sex act between two sexual partners in a steady relationship, one of whom is infected with HIV. Alternatively, the partnership over its entire duration or over the duration of the study could be defined as one potentially infective contact.

Different levels of potentially infective contacts can be defined. In the measles vaccine study in Niakhar, Senegal, four levels of exposure within a compound were defined and given a linear score. In another study of measles transmission in Niakhar, Senegal, the SARs estimated in schools, at homes and in huts differed (Cisse et al, 1999). Kendrick and Eldering (1939) differentiated definite and indefinite exposures.

### 10.6.2 Ascertainment

The method of ascertaining households for inclusion in a study is central. Households can be ascertained when a case develops within the household or a group of households can be ascertained before a case develops and followed prospectively over time. The index case of a household can be ascertained in a number of ways. A case may appear in a clinic for treatment, then the family is enrolled in the study. A case may be notified to the local authorities, and the family visited for inclusion in the study.

Prospective enrollment of households can occur in several ways. Population-based active surveillance in households at regular intervals is one method. An example is the population-based surveillance in Niakhar, Senegal. Enrollment of families prospectively, such as in the influenza studies in Tecumseh, Michigan, USA, and Seattle, Washington, USA, is another approach. In the Finnish pneumococcal carriage study, families were enrolled when the infant attended the well-baby clinics.

Ideally one would have a random sample of households in the study, whether ascertained on an index case or enrolled prospectively. Ascertainment of a household by the index case is prone to ascertainment bias. A household with a higher number of potential cases has more chance of being ascertained than a household with a smaller number. If the size of the household has an influence on the results of the analysis, then the result will be subject to ascertainment bias. It could be that households with two or more cases would more likely be ascertained than households with single cases, so that secondary attack rate would be estimated higher. However, following a large number of households prospectively could be very expensive compared to a study based on ascertaining index cases. The potential biases need to be weighed against the efficiency of the study.

In an individually randomized vaccine trial, the households of the individuals in the vaccine study can be included in a further study. If the household is included whether or not the trial participant or anyone in the household is infected, then the household is also randomized. If the household is included in a nested household study only if a case develops in the household, whether or not the first case is the vaccine trial participant or a sibling, the nested study is subject to potential selection bias.

A second issue is how are the cases within the household ascertained. If the index case is the first case in the household, then it is also the primary case. Then all further cases in contacts will be ascertained prospectively. If there are cases in a household that preceded the index case, then these cases will be ascertained retrospectively. In the PHLS pertussis vaccination study (Section 10.2.4), index cases were those cases notified to the area health authority. The household was visited, and cases within the household were ascertained both retrospectively and prospectively. Fine et al (1988) found that vaccine efficacy based on the retrospective incidence cases was lower than that based on prospective incidence cases. They proposed three possible reasons for the

observation. First, a higher number of cases in a household could result in a higher probability of ascertainment (ascertainment bias). Second, there may have been more diagnostic errors in the retrospective incidence cases (misclassification bias). False positives would reduce the efficacy estimates. The third explanation draws on the idea of the all-or-none protective effects, or at least heterogeneous protection. If the vaccine failed in some of the people, the cases in the vaccinated, unprotected people would occur early after the primary case. So the retrospective incidence cases would be enriched in vaccine failures. The vaccinated children observed prospectively would be enriched in highly protected children. Fine et al (1998) question whether retrospective incidence cases and prospective incidence cases should be lumped together in the same analysis due to potentially different sources of bias. The pertussis analysis is somewhat extreme in that a substantial portion of the retrospective incidence cases occurred more than 10 weeks before the initial visit to the household.

Onset of symptomatic disease is easier to ascertain than onset of infection. In active surveillance of symptomatic disease, surveillance could be at regular intervals and time of onset of disease retrospectively ascertained. Potential cases can be ascertained prospectively by asking family members to keep symptom diaries. When symptoms appear, they may be instructed to contact the study coordinator, or the families may be contacted regularly to check about onset of symptoms. In the carriage studies where symptoms do not occur, participants are tested at regular intervals for carriage. With infection or carriage data, the infection times between observations cannot be ascertained, but may be imputed using statistical methods.

If ascertainment of households are on an index case, then the duration of follow-up for each household needs to be determined. Household exposure studies can be used as natural challenge studies when trying to identify immunological surrogates of protection (Storsaeter et al 1998). In this situation, a decision needs to be made about the choice of timing of the immunological measurement.

### 10.6.3 Case definition

The problem of case definition is similar to other types of study designs. When households are ascertained on an index case, a different case definition is sometimes used for the secondary cases than for the index case. Retrospectively ascertained cases can often not be confirmed biologically.

### 10.6.4 Data structure

There are three basic data structures for outcomes of interest for household studies. The three are time-of-onset data, final value data, and longitudinal data. In time-of-onset data, one observes the time of onset of symptoms or infection of each of the cases in the household. In final value data, only whether

an infection or illness occurred between the beginning and end of the study period is observed for each person in the household. In longitudinal data, the members of households are followed over time and observed (sampled) repeatedly at intervals. Combinations of the types of data are possible. For example, active surveillance of households could occur at intervals. However, if a case occurs, and shows up in a clinic, then an observation occurs outside of the usual longitudinal follow-up. Time-of-onset data can be reduced to final value data for the analysis. Also, one can decide to ignore the household structure in the analysis and just analyze the data using unconditional approaches based on survival analysis or final value data.

Another important aspect of the data structure depends on the method of ascertainment. If ascertainment of a household is on an index case or index infection, then there is at least one case (infection) in each household. If ascertainment is prospective in that households are included before developing the first case, then some of the households may have zero cases. The statistical analysis may need to account for the difference in the two data structures resulting from the ascertainment method.

### 10.6.5 Assignment mechanism

We consider first that we are interested in estimating $\mathrm{VE}_S$, $\mathrm{VE}_I$, and $\mathrm{VE}_T$ from a household-based study. As is evident from equations (10.1) and (10.2), which of these efficacy parameters will be estimable depends on which secondary attack rates or transmission probabilities can be estimated. This in turn depends on who in the households are vaccinated and who are not. For example, to estimate the secondary attack rate from an infected vaccinated person to a susceptible unvaccinated person, $\mathrm{SAR}_{10}$, some of the households must have vaccinated primary cases and unvaccinated contacts. To estimate $\mathrm{SAR}_{11}$, some of the households must have vaccinated primary cases and vaccinated contacts.

Most household-based studies of vaccine efficacy up to now are observational studies or studies nested within individually randomized studies. In these studies, the allocation of vaccination within households is not under the control of the investigator. Theoretical and simulation studies have shown that to estimate $\mathrm{VE}_S$, $\mathrm{VE}_I$, and $\mathrm{VE}_T$ in the same study, discordant or individual randomization within households is better than randomization by household (Datta et al 1999, Yang et al 2006). If everyone in a household is randomized either to vaccine or control, only $\mathrm{VE}_T$ will be estimable.

Consider the four household-based influenza antiviral trials described in Section 10.3.5. The Zan II and Osel II studies both did not treat the index case, then randomized all contacts in the household to either drug or control. Thus, in both of these studies the stratified $\mathrm{AVE}_{S01/00} = 1 - \mathrm{SAR}_{01}/\mathrm{SAR}_{00}$ is estimable (AVE for antiviral efficacy). In the Osel I study, the index cases were all treated, then all household contacts randomized to either drug or control. In Osel I, the other stratified $\mathrm{AVE}_{S10/11} = 1 - \mathrm{SAR}_{11}/\mathrm{SAR}_{10}$ is estimable.

**Table 10.9.** Estimable antiviral efficacies from each of four household based, household randomized, influenza antiviral efficacy studies. $\text{AVE}_I$ is not estimale from any of the studies alone.

| | Zanamivir | | Oseltamivir | |
|---|---|---|---|---|
| | Zan I Hayden et al 2002 | Zan II Monto et al 2002 | Osel I Hayden et al 2004 | Osel II Welliver et al 2001 |
| $\text{AVE}_{S01/00} = 1 - \frac{SAR_{01}}{SAR_{00}}$ | – | $\text{AVE}_{S01/00}$ | – | $\text{AVE}_{S01/00}$ |
| $\text{AVE}_{S11/10} = 1 - \frac{SAR_{11}}{SAR_{10}}$ | – | – | $\text{AVE}_{S11/10}$ | – |
| $\text{AVE}_{I11/01} = 1 - \frac{SAR_{11}}{SAR_{01}}$ | – | – | – | – |
| $\text{AVE}_{I10/00} = 1 - \frac{SAR_{10}}{SAR_{00}}$ | – | – | – | – |
| $\text{AVE}_T = 1 - \frac{SAR_{11}}{SAR_{00}}$ | $\text{AVE}_T$ | – | – | – |

In contrast, the Zan I study randomized everyone in a household, index cases and contacts, to either drug or control. In Zan I, $\text{AVE}_T = 1 - \text{SAR}_{11}/\text{SAR}_{00}$ is estimated. Without careful examination, one might believe that all three studies were estimating the same parameter, but there could be not so subtle difference that could be important for interpreting the studies. Table 10.9 provides an overview of the efficacy estimates that can be obtained from each study. None of the four studies alone provides information to estimate $\text{AVE}_I$, the effect of the drug in reducing the infectivity of the infected index case. By combining the two oseltamivir studies or the two zanamivir studies, one can obtain estimates of $\text{AVE}_I$, though combining separate studies with other subtle design differences is not ideal.

In the pertussis vaccine study in Niakhar, there were sufficient numbers of discordant vaccinated and unvaccinated children to estimate all of the vaccine efficacies (Chapter 12.2). If it is possible control allocation of vaccination or other intervention within households at the design phase, careful consideration should be given to exactly what one would like to estimate. A study needs to be larger to get a good estimate of $\text{VE}_I$ than to estimate $\text{VE}_S$. $\text{VE}_I$ is estimated based on exposure to vaccinated compared with exposure to unvaccinated cases. If a vaccine has a strong protective effect, it may not be possible to get a good estimate of $\text{VE}_I$. However, if $\text{VE}_S$ is high, $\text{VE}_I$ has less public health importance and less influence on the results of simulation models.

## 10.7 Related designs

### 10.7.1 Case-contact design

An alternative to ascertaining clearly defined transmission units is the *case-contact* design. In the case-contact approach, an index case is identified, then

the people who have made contact with the index case are identified. For example, in tuberculosis, SARS, or HIV, through contact tracing, the people who have made contact with the infective person might be identified and their infection status ascertained. One difficulty in estimating the transmission probability from such a study is in determining the temporal order of infection in the contacts.

### 10.7.2 Cluster designs

In dengue studies, ascertainment of clusters by index cases has been used for focal mosquito control. Traditionally, a radius of 100 meters around the household of the index cases was targeted for intensive mosquito intervention. The rationale was that the usual mosquito vector of dengue virus *Aedes aegypti* has a short flight range. More recently, index cases have been used to locate clusters of people with the purpose to identify early infections in people to study the immunopathogenesis of dengue infection (Beckett et al 2005). People within a short radius of the index case are bled and followed for 14 days. The idea is that the people around an index case would be enriched for infected people compared to the general population, so that the cluster approach is more efficient than a cohort study to identify newly infected people.

Secondary attack rates in neighborhood clusters can also be used to evaluate vaccine efficacy in urban or semi-urban settings (Orenstein et al 1985). The study can be conducted by identifying neighborhood clusters, each with at least one known case. The study participants are those of the age of interest who live close to the known case. The proximity could be defined as living no more than one house away from the front doorway of the house with a case. The cluster starts at the known case in the neighborhood. The adjacent households are visited. If a case occurred in a house in the period of interest, then the houses next to it are visited until no further cases are found. Thus, all participants live within about an equal proximity to a case. The exposure is less well defined than in a household study, but perhaps better than in a population-based study. A second visit to the neighborhood will be necessary to confirm suspected cases and to detect further secondary cases.

### 10.7.3 Susceptibles exposed to infective contacts

In contrast to studies within transmission units, another study design approach to estimate the transmission probability or $\mathrm{VE}_S$ conditioning on potential exposure to infection (Yang et al 2009) is to assemble a cohort of susceptibles. The study then follows the susceptibles and collects information on their contacts with infectives or potential infectives. This type of study could be particularly useful for studies of sexually transmitted diseases or diseases transmitted by injecting drug users where contacts can be fairly easily defined. Also, the transmission probability per contact might be low. Study subjects might give information on the average number of contacts rather than

the exact number of contacts they each make per unit time. From this, the expected number of contacts during the study period can be estimated. The binomial model is probably the most commonly used model for estimating the transmission probability when susceptibles make more than one potentially infectious contact. It can take on complicated forms, depending on assumptions about variability in the transmission probability, time-varying covariates, and the amount and quality of data available. The model can be embedded in complex Markov or survival models. The principles of the binomial model were discussed in Chapter 4. The data required are infection outcome, number of potentially infective contacts, and covariate status for each person in the study. Yang et al (2009) used an errors-in-variables approach to estimate $\mathrm{VE}_S$ controlling for exposure to infection and errors in reporting number of contacts.

### 10.7.4 Augmented vaccine studies

It is possible to design studies prospectively that intentionally make use of multilevel information in estimating vaccine efficacy. One such design is the *augmented* trial design (Longini et al 1996, Datta et al 1998). In the augmented study design, individuals are recruited and possibly randomized to intervention. Then the trial can be augmented by including information on contacts and transmission units such as households or partnerships of the primary trial participants. This is one method to preserve the individual level analysis and randomization. The primary analysis can still focus on estimating $\mathrm{VE}_S$, although estimation of $\mathrm{VE}_I$ is also possible. The individual recruitment and randomization is similar to standard randomized studies that aim to estimate relative risks based on one of the unconditional measures, such as incidence rate. However, then individuals with whom the primary study participants make contact, such as in a household or partnership, are also recruited. That is, the transmission unit of the participant is recruited into the study, and augments the original primary study. The augmented participants may or may not be also randomized to intervention. Studies of vaccine efficacy based on household exposure that are nested in individually randomized clinical trials of vaccines are examples of augmented designs in which households of trial participants are recruited once a case develops in the household. The augmented study design can be thought as an extension of the idea of small transmission units within a community, as in Chapter 11, or the augmenting transmission units can be thought of as independent units, as in Chapter 12.

### 10.7.5 Mini-community designs

In a study design we call the *mini-community design*, households of individual study participants are recruited into the study, regardless of whether a case has developed in the household. The scientific goal of these types of studies is

to estimate the indirect effect of vaccination of the study participants on protecting the other household members. The goal is to estimate unconditional estimates of the type $\mathrm{VE}_{IIa}$ for indirect effects. In these studies, follow-up is over some defined period of calendar time. The goal is therefore different than in studies based on the secondary attack rates or tranmission probabilities. Similar to the community trials design, one hopes and assumes that the households are independent of one another.

If just one child in a family is in a trial, then the proportion of the family vaccinated may be too low to observe an indirect effect. That is, other siblings or household members might provide enough source of infection to mask any reduction in transmission due to the vaccinated child. If the interest is in estimating indirect effects of vaccination in families, one could consider vaccinating a larger fraction of the household. For example, in a study in South Africa, interest is on studying whether vaccinating children in the family with pneumococcal conjugate vaccine could protect HIV infected household members against pneumococcal disease. In this study, all children in some households and none in others could be vaccinated to have the maximal contrast in indirect effects.

The mini-community design is an example of a community-randomized design (Chapter 13), just that the communities are very small. The mini-community design seems particularly useful for infectious agents with a high ratio of asymptomatic infection or carriage to symptomatic disease, such as with pneumococcal bacteria. Further methodological development of the mini-community design is an open topic for future research.

## 10.8 Problems

**10.1. Different analyses**
(a) Create a hypothetical community composed of small transmission units. Assign to each individual a covariate status (0,1) and also an infection time and infection status at the end of an epidemic. Consider the various approaches for estimating the effect measures, such as the conventional secondary attack rate, the secondary attack rate and the community probability of infection simultaneously, and the simple incidence proportion.
(b) How do the data being used for each approach differ? What parameters can be estimated? What is the interpretation of the measures under each approach?

**10.2. Ignoring household structure**
(a) Consider the data in Table 10.4. Ignoring the household structure, compute the attack rate for each different household size and the study population as a whole.
(b) Is there any trend in the attack rates by size of household? Would you expect one? Why or why not?

**10.3. Exposure in measles study**

(a) Consider the data in Table 10.8. Define the rate of exposure to measles as the number of children exposed divided by the number at risk on January 1, 1990. Compute the rate of exposure for the three vaccine groups and the unvaccinated group. Are there any differencs in the exposure rates among the groups?

# 11

# Analysis of Households in Communities

## 11.1 Overview

In this chapter, we consider analyses that assume that the households or other transmission units are nested in a community. Community-acquired infection serves as a source of initial infection within households as well as possible further cases in the household. Infected household members can infect others in the household. To start, we discuss general aspects of these models. All models in this chapter are variants of the basic models presented in this section. They use different data structures, assumptions, and methods of estimation, but the underlying parameters are similar. The data can be final-value data, time-to-event data, or longitudinal (panel) data.

Each model has two general types of parameters, one for infection from the community, the other for transmission from an infective to a susceptible within the household. The first is an unconditional parameter, that is, it does not condition on exposure to infection, the second a conditional parameter. The models can be formulated in discrete time or continuous time. For some data structures, such as contact data on sexual contacts, contacts can be substituted for time. Models formulated in discrete time have a parameter for the probability of infection from the community per unit time and a parameter for the probability of transmission from an infective to a susceptible within the household per unit time. Continuous time models have analogous rate parameters. One parameter describes the rate of community-acquired infection, the other the rate of transmission from an infective to a susceptible within a household. Both continuous and discrete time parameters can be transformed into the probability of acquiring infection from the community over the period of time of interest, called the community probability of infection, CPI, and the secondary attack rate within the household, SAR. If an estimate of prevalence of infection in the population is available, then the transmission probability in the community at large can also be estimated (Hudgens, et al, 2001).

The approaches in this chapter are not used as often as the conventional secondary attack rate (Chapter 12) for vaccine evaluation, but they could be. Not much standard software is available for the methods in this chapter. The approaches generally require statistical knowledge and computer programming skills.

### Discrete-time model

Consider a study from time period 0 to time period $T$. Let $a$ be the probability a susceptible household member becomes infected from the community in one time unit. Let $b = 1 - a$ be the corresponding escape probability. Then the probability of escaping infection from the community over the $T$ time periods is $B = b^T = (1 - a)^T$, and the community probability of infection

$$\text{CPI} = 1 - B = 1 - b^T = 1 - (1 - a)^T. \tag{11.1}$$

Let $q = 1 - p$ be the probability of escaping infectious contact in a household within one time unit. Then if a person is infectious for $T_I$ time units, the probability of escaping infection from an infective within a household is $Q = q^{T_I} = (1 - p)^{T_I}$, and the secondary attack rate

$$\text{SAR} = 1 - Q = 1 - (1 - p)^{T_I}. \tag{11.2}$$

### Continuous-time model

In the continuous time model, a parameter $\alpha$ denotes the instantaneous risk of infection from the community and a parameter $\beta$ denotes the instantaneous risk of infection from an infective in the household. In the simplest form, if the study duration is from time 0 to time $T$ and the duration of infectiousness is $T_I$, then

$$\begin{aligned} \text{CPI} &= 1 - \exp(-\alpha T), \\ \text{SAR} &= 1 - \exp(-\beta T_I). \end{aligned} \tag{11.3}$$

### Vaccine effects and other covariates

Vaccination status and other covariates can be easily entered into the models. Either separate values of each parameter can be estimated for each category or parameters representing the effects of covariates can be included in the model. Typically the parameter $\theta$ denotes the relative susceptibility of a vaccinated compared to an unvaccinated person, so that $\text{VE}_S = 1 - \theta$. Similarly, the parameter $\phi$ denots the relative infectiousness per contact of a vaccinated compared with an unvaccinated person, so that $\text{VE}_I = 1 - \phi$. If $p$ is the per time unit (or per contact) transmission probability between two unvaccinated

people in a household, then $\theta p$ is the per time unit (or per contact) transmission probability to a vaccinated susceptible from an unvaccinated infective. The secondary attack from an unvaccinated infective individual to a vaccinate susceptible is $\mathrm{SAR}_{01} = 1 - (1 - \theta p)^{T_I}$.

If only the infective person is vaccinated, then $\mathrm{SAR}_{10} = 1 - (1 - \phi p)^{T_I}$. If both people are vaccinated, then $\mathrm{SAR}_{11} = 1 - (1 - \theta\phi p)^{T_I}$. This latter model assumes that the vaccine's effects on infectiousness and susceptibility in reducing the transmission probability are independent and multiplicative. Alternatively, one could use another parameter $\psi$ to denote the vaccine's effect on the transmission probability if both the infective and the susceptible in the contact. are vaccinated, so that $\mathrm{SAR}_{11} = 1 - (1 - \psi p)^{T_I}$. The vaccine parameters enter similarly into the continuous time models. For example, $\mathrm{SAR}_{11} = 1 - \exp(-\theta\phi\beta T_I)$.

The CPI involves only the susceptibles directly. One could introduce another parameter $\theta_c$ denoting a different effect of vaccination on reducing susceptibility to infection from the community. However, introducing more parameters into a model sometimes cannot be supported by the amount of data available. Often the assumption is made that the effect of vaccination on protecting against infection from the community and from an infective in a household are the same, so that just one parameter $\theta$ is estimated. This is a strong biological assumption, since exposure within a household could be more intense (Fine et al 1988).

Other covariates such as age can be entered similarly into the model. Commonly, child- and adult-specific transmission probabilities or rates are estimated Covariates such as vaccination or treatment status can change over time in models that incorporate time. The parameters $a$ and $p$, or $\alpha$ and $\beta$ can also be time-dependent. Information about the prevalence of infection can be used to estimate the community probability or rate of infection as varying over time, so that $a$ or $\alpha$ can be functions of time. Infectiousness of an infective within a household can vary with time after being infected, so $p$ or $\beta$ can be functions of time after being infected.

Note that if we define $\mathrm{VE}_{S,p} = 1 - \theta$, this will not necessarily equal

$$VE_{S,SAR} = 1 - \frac{SAR_{01}}{SAR_{00}} = 1 - \frac{1 - (1 - \theta p)^{T_I}}{1 - (1 - p)^{T_I}}. \tag{11.4}$$

See Problem 11.1.

### Estimation

Estimation is most often in a likelihood or Bayesian framework. The likelihood component of the Bayesian model is sometimes exactly the same as that in the likelihood framework. In general, though, a Bayesian framework that uses Markov chain Monte Carlo (MCMC) methods for estimation allows for relaxation of some assumptions of a straight likelihood approach, though sometimes at the expense of making other assumptions.

In the next two sections, we consider models for final value data and time-of-onset data for diseases in which individuals acquire infection or disease just once over the course of the study. These correspond to either the SEIR or SIR models. In Section 11.4 we present models for longitudinal data for infections with repeated acquisition and clearance of infection, such as in pneumococcal carriage studies. These correspond to SIS models.

## 11.2 Final-value data

### 11.2.1 Data structure

The data required are the number of susceptibles in each household at the beginning of the observation period and the number of infections that occurred in each household by the time the observation period is over. Household final-value data for influenza infection are in Table 10.4. Data further categorized by covariates, such as vaccination, antibody titers (Table 10.5), or age, allow the estimation of the effects of covariates.

Assume that observations are made on infections in a community, starting in time period $t = 0$ and ending in time period $t = T$. This period could correspond to an epidemic season or some other period of epidemiological interest. The main criterion is that all, or nearly all, of the outbreaks in the sample of households should essentially have run their course within $[0, T]$. The final value data on $n$ households are observed, where $a_{jk}$ = observed number of households with $k$ original susceptibles of which $j$ become infected, $k = 1, 2, \ldots, K$ and $j = 0, 1, \ldots, j$, where $\sum_k \sum_j a_{jk} = n$. For example, $a_{13} = 4$ means that there are four households with three household members in which one person became infected. This analysis requires biologic confirmation of susceptibility before and infection after. Categorical covariates such as vaccination status or age could also be observed. For example, people could be either unvaccinated or vaccinated, denoted 1 and 2 respectively. Then $a_{(j_1,j_2)(k_1,k_2)}$ = observed number of households in which $(j_1, j_2)$ of $(k_1, k_2)$ susceptibles in each household become infected.

### 11.2.2 Discrete-time model

Longini and Koopman (1982) present a model for the distribution of the total number of cases in households from a homogeneous community and use a maximum likelihood approach for estimation. To derive the final-size distribution of household infections, three key assumptions are made: (1) sources of infection from the community are distributed homogeneously; (2) household members mix at random within the household; (3) each household member can be infected either from within the household or from the community.

Infection from the community is modeled by defining $a_t$ as the probability a susceptible household member becomes infected from the community in

time period $t$, and $b_t = 1 - a_t$ is the corresponding escape probability. Define $B$ as the probability that a susceptible individual is not infected from the community during the period of observation. A general expression for $B$ is

$$B = \prod_{t=0}^{T} f(b_t), \tag{11.5}$$

where $f(\cdot)$ is bounded function describing infection rates in the community. A simple form for $f(\cdot)$ is $f(b_t) = b_t$.

Now consider the effect of secondary spread within a household following introductions from the community. An individual is infected in time period $t_0$ and will pass through a series of stages in time periods $t_1, t_2, \ldots$, until becoming immune. Define $p_t$ as probability that an infective who was infected at time $t = t_0$ will make infectious contact in the household with another individual in time period $t$. Then $\{p_t\}$ describes pattern of infectiousness over time. The structure of $\{p_t\}$ is

$$\begin{array}{lll} p_t = 0 & \text{when} \;\; t_0 \le t \le t_l, & \text{the latent period} \\ p_t > 0 & \text{when} \;\; t_{l+1} \le t \le t_m, & \text{the infectious period} \\ p_t = 0 & \text{when} \;\; t_{m+1} \le t < t_\infty, & \text{the immune period.} \end{array}$$

Let $q_t = 1 - p_t$ be the daily probability of escaping infectious contact. Then if there is an infected individual in the household who became infected at time $t = t_0$, let $Q_{t_r}$ be the probability that a susceptible individual has escaped infection within the household at time $t_r$, $t_0 \le t_r < t_{m+1}$. The probability $Q$ that the susceptible individual escapes infectious contact from the infective during his entire period of infectiousness is

$$Q = \prod_{t=t_0}^{t_m} q_t = Q_{t_m}$$

As described in Section 11.1, the secondary attack rate $\mathrm{SAR} = 1 - Q$, and the community probability of infection $\mathrm{CPI} = 1 - B$.

### Final-size distribution

Assume all households under consideration are free of infected members at the beginning and end of the period of observation. Let $\Pr(j|k)$ be the probability that $j$ of $k$ initial susceptibles within a household are infected the course during the course of the epidemic. Write $m_{jk} = \Pr(j|k)$ to simplify notation.

When $k = 1$, it follows from the above assumptions that $m_{01} = B$ and $m_{11} = 1 - B$. When $k = 2$, $m_{02} = B^2$. For $m_{12}$, there are two possible ways it can occur. Either the first susceptible individual becomes infected with probability $B$, and the second susceptible escapes infection from both the community and the infective in the household, or the first susceptible

escapes infection from both sources and the second becomes infected in the community. Thus

$$m_{12} = 2(1-B)BQ = 2m_{11}BQ.$$

For $m_{22}$, similarly

$$\begin{aligned} m_{22} &= 2(1-B)(1-Q)B + (1-B)^2 \\ &= 1 - m_{02} - m_{12}, \end{aligned}$$

since the probabilities must sum to one. In general, there are $\binom{k}{j}$ ways to get $j$ infected individuals from $k$ originally susceptible ones. The general expression for $m_{jk}$ is

$$\begin{aligned} m_{jk} &= \binom{k}{j} m_{jj} B^{k-j} Q^{j(k-j)}, \quad j < k, \\ m_{kk} &= 1 - \sum_{j=0}^{k-1} m_{jk}. \end{aligned} \tag{11.6}$$

The density function (11.6) provides the final-size distribution for the modified Reed-Frost model of Sugiyama(1960).

If it is assumed that there is spread only within the household, then $B = 1$. If there are initially $i$ infectives within the household, then equation (11.6) becomes

$$m_{jk} = \binom{k}{j} m_{jj} Q^{i+j(k-j)}, \quad j < k,$$

and $i + j$ is the final number of infectives in the household, equivalent to equation (4.8).

When $Q = 1$, there is no spread in households, and the disease in question is presumably not infectious. Then (11.6) reduces to the binomial distribution:

$$m_{jk} = \binom{k}{j} (1-B)^j B^{k-j}, \quad j \le k. \tag{11.7}$$

This is the distribution of infection in household we would expect if the disease were not contagious, and we would analyze final attack rate data in a community of households. If household structure is ignored or there is no household spread, then CPI $= 1 - B$ is the incidence proportion, or attack rate.

In some cases, the zero class $a_{0k}$ is not present. This occurs when households are surveyed only after an initial infective has appeared. Then the zero-truncated distribution is used. The general expression for $m_{jk}$ is then

$$m_{jk} = \binom{k}{j} m_{jj} B^{k-j} Q^{j(k-j)} / (1 - B^k), \quad j < k. \tag{11.8}$$

**Table 11.1.** Observed and expected distributions of Asian influenza data (Sugiyama, 1960) in households of size three as analyzed by Longini and Koopman (1982)

| Number of cases | Observed number of households | Expected number of households |
|---|---|---|
| 0 | 29 | 29.17 |
| 1 | 9 | 7.87 |
| 2 | 2 | 3.62 |
| 3 | 2 | 1.34 |
| Total | 42 | 42.00 |

### Likelihood Estimation

The parameters $Q$ and $B$ can be estimated by maximum likelihood. The likelihood function is

$$L(Q,B) = \prod_k \prod_j m_{jk}^{a_{jk}}.$$

The explicit form of the log likelihood function from (11.6) is

$$\ln L = c + \sum_k \sum_j a_{jk}\{\ln m_{jj} + (k-j)\ln B + j(k - j\ln Q\}).$$

The ML estimates $\hat{Q}$ and $\hat{B}$ are solutions of the score functions that can be solved iteratively by the method of scoring. The information matrix provides variances and covariances. Estimates from the data provide starting values. In the truncated case, the ML procedure can proceed, using a different approach to get initial guesses. Becker (1989), pp. 182–193 discusses a similar model with an approximate approach for estimation. Becker (1989) and Haber, et al.(1988) use a generalized linear model.

### Analysis of data from an Asian influenza epidemic

Table 11.1 presents data from an Asian influenza epidemic from households with three initially susceptible people. The data are the number of households that had either 0, 1, 2, or all 3 people infected by the end of the epidemic.

Using model (11.6), $\hat{B} = 0.856$, $\text{var}(\hat{B} = 0.0009)$ and $\hat{Q} = 0.834$, $\text{var}(\hat{Q} = 0.0063)$. Thus, the estimated probability of a susceptible individual being infected by an infective in his household during the course of his infectious period is $\widehat{SAR} = 1 - \hat{Q} = 0.166$. Assuming that the latent period $l = 2$ days and infectious period $T_I$ is 4 days, and $p_t = p$ for $t = 3, 4, 5, 6$, then the estimated daily probability of escaping infection in the household

**Table 11.2.** Comparison of CPI and SAR from the influenza A(H3N2) epidemic seasons 1977-1978 and 1980-1981 combined, In Tecumseh, Michigan, stratified by age group and pre-season antibody titer (from Longini et al 1988)

| | | Pre-season antibody titer $(1:x)$ | |
|---|---|---|---|
| Age | | Low level $(x < 8)$ | High level $(8 \leq x \leq 64)$ |
| Children | $\widehat{\text{CPI}}$ | $0.231 \pm 0.032$ | $0.094 \pm 0.028$ |
| (0–17 years) | $\widehat{\text{SAR}}$ | $36.6 \pm 6.2$ | $3.4 \pm 4.7$ |
| Adults | $\widehat{\text{CPI}}$ | $0.131 \pm 0.018$ | $0.089 \pm 0.015$ |
| (18+ years) | $\widehat{\text{SAR}}$ | $18.2 \pm 4.4$ | $1.6 \pm 3.7$ |

is $\hat{q} = Q^{1/T_I} = 0.834^{0.25} = 0.956$. The daily probability of infection in a household $p = 1 - q = 0.044$.

The estimated $\widehat{\text{CPI}} = 1 - \hat{B} = 0.144$. The approximate percentage of cases from the community can be calculated by setting $Q = 1$. Then from (11.7), if there no spread within the household, the expected number of cases would be

$$nk(1 - \hat{B}) = 14.4,$$

but the total number of observed cases allowing spread within the household ($\hat{Q} = 0.834$) is 19. Hence, approximately 75% of total cases were due to infection from the community.

To further illustrate the role of the mixing assumptions in this model, we can estimate the usual attack rate from these data by simply ignoring the household structure. Suppose we do not have information on households. There are 42 households with three people each, so the total population is 126 people. From Table 11.1, 19 people became infected. The attack rate is AR= 19/126 = 0.151. The attack rate is interpreted as the probability of becoming infected without any further assumptions about the dynamics of interaction in the community or households. The estimated $AR$ is higher than the estimate of community probability of infection, CPI. The simple AR is higher than the CPI because the AR includes the portion of the infected individuals who, under the model that included the SAR, were estimated to have been infected within households. This example illustrates the importance of considering the mixing assumptions within a population when developing models for estimating meaningful population parameters in infectious disease epidemiology.

### Extension to covariates

Longini et al (1988) extended the model to include categorical covariates such as age group, antibody level, or vaccine status. Assuming that people are equally infectious regardless of stratum, they computed the SAR and CPI

for children and adults stratified further by the level of pre-season antibody. The summary data are in Table 10.5 and the estimates are in Table 11.2. An antibody efficacy similar to vaccine efficacy can be computed as

$$\text{Antibody efficacy} = 1 - \frac{\text{Risk(high antibody titer)}}{\text{Risk(low antibody titer)}}, \tag{11.9}$$

using the estimates of SAR and CPI in Table 11.2 and AR in Table 10.5 as the measures of risk (See Problem 11.2)

**Using Markov chain Monte Carlo methods**

O'Neill et al (2000) used Bayesian inference to estimate the probability of escape from infection in the community and from an infective in a household. The likelihood part of their Bayesian model was the same as in Longini and Koopman (1982), with a different recursive approach to obtain the final-size distribution. Using uniform prior distributions for the parameters $B$ and $Q$, the posterior density should be equivalent to the likelihood. A Metropolis-Hastings algorithm was used for Bayesian inference. Figure 11.1 shows the joint posterior distribution of the two parameters of interest for the Tecumseh data in Table 10.4, where in the figure $q_c = B = 1 - \text{CPI}$, $q_h = Q = 1 - \text{SAR}$. Applying a simple numerical maximization technique to the MCMC output yielded estimates for the Tecumseh data that were virtually identical to those in Addy et al (1991) (Section 11.2.3). Both O'Neill et al (2000) and Addy et al (1991) obtained $q_c = Q = 0.8677$. O'Neill et al (2000) obtained $q_h = B = 0.8408$ and Addy et al (1991) give $q_h = B = 0.8406$.

### 11.2.3 Generalized stochastic model

A stochastic infectious disease model was developed by Ball (1986) in which the distribution of the length of the infectious period is allowed to have any distribution that can be described by a Laplace transformation. Addy, et al (1991) extended this model to allow for infection from an unspecified source in the community or transmission within a household. The model allows for discrete covariates, such as age group or vaccine status, for heterogeneous susceptibility and infectivity. The model for the probability of escaping infection in the community, $B = 1 - CPI$ is formulated in discrete time. However, the transmission parameter to be estimated is $\beta_{ik}$, the rate at which a susceptible of type $i$ has contact with infective of type $k$. The final size distribution is found recursively. In the special case of a constant infectious period, the final size distribution is the same as that described above.

If $T_I$ is a constant duration of infectivity and $\beta$ does not vary, then $\exp(-\beta T_I)$ is the probability of escaping infection by an infective. Then SAR$= 1 - \exp(-\beta T_I)$. When $T_I$ is variable, the SAR is calculated by taking the expectation, and SAR$= 1 - E\{\exp(-\beta T_I)\} = 1 - \phi(\beta)$, where $\phi(\cdot)$

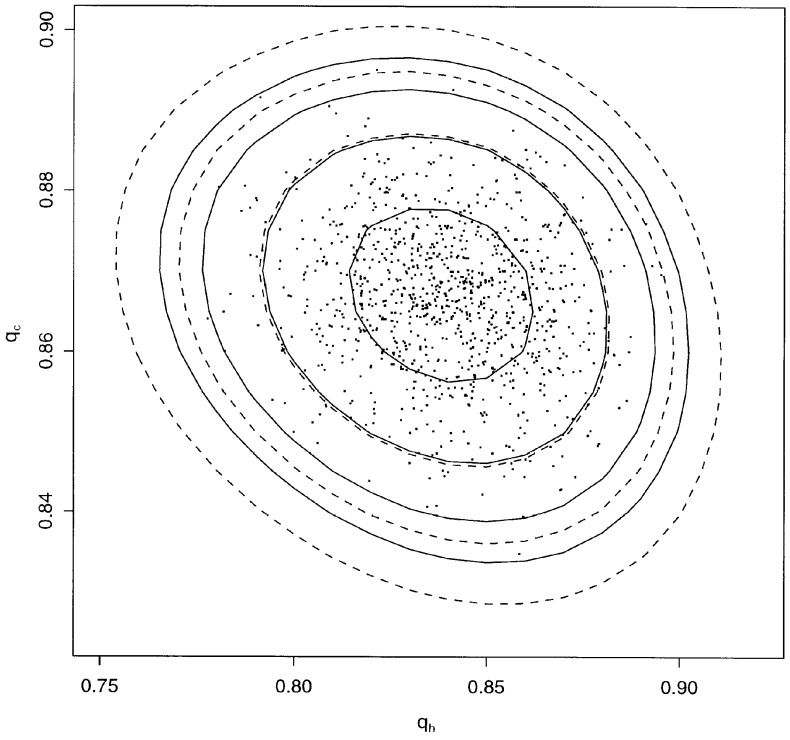


**Fig. 11.1.** Joint posterior distribution, $q_c = B = 1 - CPI$, $q_h = Q = 1 - SAR$, analyzing Tecumseh influenza data from Addy et al (1991). MCMC sample values (1000 values, at sampling interval 10): ———, contour lines surrounding highest posterior density credible intervals at 50%, 90%, 99% and 99.9% levels; - - - -, posterior probability density function values of 10%, 1% and 0.1% of its maximum (from O'Neill et al 2000).

is the Laplace transform of the length of the infectious period. The standard error of the SAR is calculated using the delta method on the Laplace transform, if $T_I$ is variable. Estimates were obtained using maximum likelihood. The results of the analysis of the Tecumseh influenza data in Table 10.4 is in Table 11.3 for constant $T_I$ assuming homogeneity and in the second analysis allowing for age-group specific transmission rates.

### 11.2.4 Other final-value analyses

Becker and Angulo (1981) use household data that includes smallpox vaccination status from an epidemic of variola minor, a mild form of smallpox (Angulo 1976) to estimate the protective effects of smallpox vaccination. Becker

**Table 11.3.** Maximum likelihood estimates and standard errors for parameters of model of influenza A(H3N2) infections in 1977-1978 and 1980-1981 combined epidemics in Tecumseh, Michigan assuming contant period of infectiousness $T_I$ = 4.1 days (from Addy et al 1991). For transmission, the first subscript refers to susceptibility, the second to the infective. SAR given in %.

| | Estimate | Transformation |
|---|---|---|
| Homogeneity: | | |
| | $\beta = 0.0423 \pm 0.0061$ | SAR$= 15.9 \pm 2.1$ |
| | $B = 0.8677 \pm 0.0097$ | CPI$= 0.1323 \pm 0.0097$ |
| Log likelihood | $= -532.974$ | |
| Child= 1, Adult= 2 | | |
| | $\beta_{11} = 0.0805 \pm 0.0208$ | $\text{SAR}_{11} = 28.1 \pm 6.1$ |
| | $\beta_{12} = 0.0354 \pm 0.0291$ | $\text{SAR}_{12} = 13.5 \pm 10.3$ |
| | $\beta_{21} = 0.0268 \pm 0.0135$ | $\text{SAR}_{21} = 10.4 \pm 5.0$ |
| | $\beta_{22} = 0.0401 \pm 0.0127$ | $\text{SAR}_{22} = 15.2 \pm 4.4$ |
| | $B_1 = 0.8184 \pm 0.0254$ | $\text{CPI}_1 = 0.1816 \pm 0.0254$ |
| | $B_2 = 0.8897 \pm 0.0128$ | $\text{CPI}_2 = 0.1103 \pm 0.0128$ |
| Log likelihood | $= -522.333$ | |

et al (2003) reanalyze those data and use Bayesian inference to estimate the probability that smallpox vaccination is completely protective and the relative susceptibility and infectiousness in those only partially protected. Magder and Brookmeyer (1993) use a generalized linear model and EM algorithm to estimate the community probability of transmission and transmission parameters on the example of intravenous drug users and HIV. In a study of dengue transmission, Dantes et al (1988) used the model to estimate the relative risk of transmission at both the individual and the household level.

## 11.3 Time-of-onset Data

### 11.3.1 Likelihood approach

Yang et al (2006) developed likelihood methods to estimate the prophylactic effects of interventions in households using time-of-onset data. They compared randomization schemes and prospective versus retrospective ascertainment. The methods were an extension of the discrete time data approach (Rampey et al 1992). The methods were motivated by the influenza antiviral household studies in Table 10.7 and use to analyze the two oseltamivir trials. They are also applicable to vaccines and other prophylactic agents. The model assumes that the distributions of the latent and the infectious period (Table 11.4) and does not take asymptomatic infection into account. The latent period is assumed to be of the same duration as the incubation period, so that a person is assumed infectious once symptoms develop.

**Table 11.4.** Empirical cumulative distributions of the latent period and the infectious period for influenza (Elveback et al., 1976).

| Latent Period | | Infectious Period | |
|---|---|---|---|
| Duration (days) | Cumulative Probability | Duration (days) | Cumulative Probability |
| 1 | 0.2 | 3 | 0.3 |
| 2 | 0.8 | 4 | 0.7 |
| 3 | 1.0 | 5 | 0.9 |
| | | 6 | 1.0 |

Let the trial start on day 1 and end on day $T$. The simplest data for each participant are the first date with symptoms of the disease of interest, the assigned treatment, and the treatment period. Let $p$ be the transmission probability per daily contact within the household between a susceptible person and an infective person if both have not received the treatment. Let $b$ be the daily probability that a susceptible, untreated person is infected by a source of infection from the community.

Analogous to vaccine efficacies, the antiviral efficacies can be estimated. Let $\text{AVE}_S = 1-\theta$, where $\theta p$ is the reduced transmission probability resulting in illness if the susceptible person is taking an antiviral agent and exposed to an untreated infected person in the household. The model assumes that efficacy is the same for contacts outside the household, i.e., the reduced transmission probability resulting in illness for a person taking an antiviral agent is $\theta b$. $\text{AVE}_I = 1-\phi$, where $\phi p$ is the reduced transmission probability if the infective person is treated. $\text{AVE}_T$ is the total effect on transmission when both people in a transmission pair are treated. The analysis considered the two different assumptions of independence and multiplicativity of $\theta$ and $\phi$ as well as that a separate parameter $\psi$ be estimated if both the infective and the susceptible in a contact received treatment. Here we present just the former.

### Notation and escape probabilities

Let $\tilde{t}_i$ denote the day of illness onset for an infected person $i$. Let $r_i(t) = (0$ untreated, 1 treated) indicate the treatment status of person $i$ on day $t$. Let the function $f(t|\tilde{t}_j)$ denote the probability that person $j$ is infectious on day $t$ given the day of illness onset $\tilde{t}_j$. Assuming independence between $\theta$ and $\phi$, the probability that a susceptible person $i$ escapes infection by an infective family member $j$ on day $t$ is given by

$$q_{ij}(t) = 1 - \theta^{r_i(t)}\phi^{r_j(t)} p f(t|\tilde{t}_j), \tag{11.9}$$

Let $D_i$ denote the set of people in the same household with person $i$. Then

$$e_i(t) = (1 - \theta^{r_i(t)} b) \prod_{j \in D_i} q_{ij}(t) \quad \text{and} \quad Q_i(t) = \prod_{\tau=1}^{t} e_i(\tau) \tag{11.9}$$

**Table 11.5.** Maximum likelihood estimates by age (1-17 vs 18+) for pooled oseltamivir trials conducted in 1998-1999 and 2000-2001, North America and Europe (Yang et al 2006).

| Assumption $\psi = \theta\phi$ | Parameter | MLE | 95% C.I. | SAR | Estimate | 95% C.I. |
|---|---|---|---|---|---|---|
| Yes | $b_c$† | 0.0023 | (0.0015, 0.0035) | | | |
| | $b_a$ | 0.00055 | (0.0003, 0.001) | | | |
| | $p_{cc}$ | 0.038 | (0.023, 0.063) | $SAR_{cc}$‡ | 0.15 | (0.074, 0.21) |
| | $p_{ca}$ | 0.012 | (0.007, 0.021) | $SAR_{ca}$ | 0.049 | (0.021, 0.075) |
| | $p_{ac}$ | 0.018 | (0.008, 0.040) | $SAR_{ac}$ | 0.071 | (0.014, 0.13) |
| | $p_{aa}$ | 0.022 | (0.014, 0.034) | $SAR_{aa}$ | 0.086 | (0.047, 0.12) |
| | $AVE_S$ | 0.85 | (0.52, 0.95) | | | |
| | $AVE_I$ | 0.66 | (-0.10, 0.89) | | | |
| | $AVE_T$ | 0.95 | (0.77, 0.99) | | | |

†, ‡ Subscription $c$ denotes child (1-17), $a$ denotes adult (18+), and $ca$ denotes child-to-adult transmission.

‡ $SAR_{vu}$ is based on the average 4.1 days of infectious period, i.e., $SAR_{vu} = 1 - (1 - p_{vu})^{4.1}$.

are the escape probabilities for person $i$ on day $t$ and up to day $t$, respectively, The probability that person $i$ is infected on day $t$ is given by

$$Z_i(t) = Q_i(t-1)(1 - e_i(t)). \tag{11.10}$$

Allowance must be made that we do not observe the exact infection times, but just the onset of illness. Assuming the duration of the latent (and incubation) period is known, then it is possible to compute the maximum and minimum duration of the latent period. Let $\underline{t}_i$ and $\bar{t}_i$ be the earliest and latest potential infection day for person $i$. Let $g(\tilde{t}_i|t)$ be the probability of illness on day $\tilde{t}$, given infection on day $t$. Then the contribution to the likelihood of person $i$ is

$$L_i(b, p, \theta, \phi, \psi|\tilde{t}_j, j \in D_i) = \begin{cases} Q_i(T) & \text{if } i \text{ is not infected} \\ \sum_{t=\underline{t}_i}^{\bar{t}_i} g(\tilde{t}_i|t) Z_i(t) & \text{otherwise} \end{cases} \tag{11.11}$$

MLEs can be obtained using usual methods. Results are in Table 11.5.

### 11.3.2 Bayesian latent variable approach

Cauchemez et al (2004) adapted the Bayesian hierarchical model of Auranen et al (2000) (Section 11.4.1) for influenza household studies. The essential difference is that unlike in the pneumococcal carriage studies, individuals can have only one episode of influenza. That is, the model for influenza assumed an SIR model, rather than an SIS as in Section 11.4.1. A main difference in this approach compared to the likelihood method in Section 11.3.1 is that the

distribution of the duration of the infectious period is estimated whereas in the previous model it was assumed known. In this latent variable model, it is assumed that individuals became infectious as soon as infected, that is, there is no latent period. In the likelihood model, the latent period distribution was assumed known and greater than 0. This model assumes there are no inapparent infections, similar to the likelihood model above that does not take asymptomatic infections into account.

The study motivating the analysis is described in Chapter 10.3.4. The data were times of ascertainment of culture-confirmed index cases in physicians' practices and the time of onset of clinical influenza (not biologically confirmed) within ascertained households. Similar to Auranen et al (2000), the unobserved start and end of the infectious period for each case of influenza were imputed using a data augmentation algorithm. Let $I(t)$ be the collection of infectives in a household. Let $\alpha_i$ be the rate of transmission from the community for individual $i$. Let $\epsilon_i\beta_j/n$ be the rate of transmission from an infective $j$ to a susceptible $i$ in a household of size $n$. For an individual $i$ susceptible just before $t$, the rate of transmission is

$$\lambda_i(t) = \alpha_i + \epsilon_i \sum_{j \in I(t)} \beta_j/n, \tag{11.12}$$

similar to model (11.15). The distribution of the infectious period was assumed to be a gamma distribution, for infective $i$ with mean $\mu_i$, standard deviation $\sigma_i$, and density $d_{\mu_i,\delta_i}$. The distribution of the infectious period was estimated from augmented data for the unobserved dates of the start and end of the infectious period. A strong gamma prior was used for the distribution of the infectious period.

The community probability of infection (CPI) is defined as the probability that participant $i$ would be infected from the community during the 15 day follow-up period of the household,

$$\mathrm{CPI}_i = 1 - \exp(-15\alpha_i).$$

The secondary attack rate, SAR, defined as the probability that an infective $j$ infects susceptible $i$ in a household of size $n$ is

$$\mathrm{SAR}_{ij}(n) = 1 - \int_0^{+\infty} \exp\left(-\epsilon\frac{\beta_j}{n}\right) \mathrm{d}_{\mu_j,\delta_j}(t)\mathrm{d}t \tag{11.13}$$

The interest was in exploring the role of children in the infection process, so parameters were stratified by children under 15 and adults.

The mean duration of influenza infectious period was estimated at 3.8 days (95% CI [3.1,4.6]) with standard deviation of 2.0 days (95% CI [1.1,2.8]) for a 95% credible interval for the infectious period of (0.8,8.6) days for influenza. The overall household SAR attack for clinical influenza decreased from 0.43 (95% CI 0.39, 0.48) for households of size 2 to 0.21 (95% CI 0.18,024). This

result could be partly model dependent. The overall CPI over the 15 days of followup was 0.08 (95% CI 0.04, 0.12). They found similar age-specific trends in transmission parameters as others.

### 11.3.3 Other time-of-onset analyses

A data-augmentation method was developed to analyze the two zanamivir trials in Table 10.7 (Yang et al 2007). Similar models were used as the basis for a resampling-based test to detect person-to-person transmission of an infectious disease (Yang et al 2007a) and to detect human-to-human transmission of Avian A(H5N1) influenza (Yang et al 2007c). The software TranStat is publicly available for this purpose.

## 11.4 Longitudinal Data

Many bacterial infections are characterized by repeated acquisition and clearance of infection. Asymptomatic carriage of pneumococcal bacteria and *H. influenza b* (Hib) bacteria are two examples. To estimate the acquisition and clearance rates, one needs longitudinal data where a collection of individuals in households is sampled over time at a number of time points. Several field studies that gathered longitudinal data from repeated sampling within families or schools are described in Chapter 10.5. The problem with longitudinal data of asymptomatic carriage is that neither the acquisition or clearance times are observed. Different approaches to statistical models can deal with this problem making a variety of assumptions.

Motivated by two studies of Hib carriage, Auranen et al (1996) proposed the use of a susceptible-infected-susceptible (SIS) model for estimating the acquisition and clearance parameters. At any given time, individuals can be in either the non-carrier, susceptible state S, or the asymptomatic, infectious carrier state C. People can make the transition between states, $S \to C$ and $C \to S$. The mini-epidemic in the family is represented by an individual-based stationary Markov model that describes the dynamics of infection in the family. The approach is also appropriate for use with pneumococcal carriage studies. Other problems with pneumococcal studies include combining multiple serotype data, missing data, competition, and errors in diagnosis.

The key parameters in the model are the hazard rates of transitions $S \to C$ and $C \to S$. Let $C(t)$ be the number of carriers in the family at time $t$. The basic model before allowing for covariates, for the transition $S \to C$ is a combination of the effective contact rate within the familly $\beta$ and the rate of infection from the community $\alpha$. The rate of transition $S \to C$

$$\lambda^{(i)}(t) = \alpha^{(i)} + \beta^{(i)} C^{(i)}(t). \tag{11.14}$$

One can also make the transitions age-group dependent, such as children and adults. The constant rate of transition $C \to S$ is denoted by $\mu^{(i)}$. The spread

of carriage within a family is modeled by a Markov process. The state at time $t$ is the combined state of the individuals in the family. The possible transitions within a family in any time interval can be described by a matrix of transition probabilities. However, the matrix is of large dimension leading to possible computational difficulties. This approach is discussed in more detail in Section 11.4.2.

A second approach was suggested by Auranen et al (2000) that makes use of Bayesian data augmentation. For each individual, the unobserved times of acquiring and clearing carriage are included in the model as latent, unobservable variables. If the acquisition and clearance times of each individual were known, the conditional likelihood of the data is simple. Markov chain Monte Carlo methods are used to augment the data with the acquisition and clearance rates. The approach considers histories of infection events separately for each individual.

Pneumococcal carriage studies and the models in this section fit well within the context of a general framework of estimating vaccine effects. Vaccine parameters for susceptibility and infectiousness, also vaccine-dependent clearance rates, could be added to the models. The question of how to use these approaches in pivotal trials of vaccines for licensure is open.

### 11.4.1 Bayesian Latent Variable Approach

Auranen et al (2000) modeled the sequences of binary observations on Pnc carriage by constructing latent point processes of acquiring and clearing carriage. The study motivating this analysis is described in Chapter 10.5.1. The model allows for carriage of different serotypes $s$. The study identified seven Pnc serotypes that were going to be included in the planned vaccine. The analysis was confined to the three most prevalent types, 6B, 19F, 23F. The data for children under 5 years old are summarized in Table 11.6. The observed numbers of changes in the carriage status over the observation intervals are presented. The data are stratified according to age class, 0 to 2 years, 2 years to 5 years. They are also stratified by the background carriage in the family at the start of the observation period. The table summarizes the data over the three serotypes. Pnc carriage was almost always clustered by serotype within a family.

The hierarchical model has three stages: 1) the observation model, 2) the transmission model, and 3) the prior model. In the observation model, the augmented data need to be consistent with the observed data. Any observation of carriage of that serotype needs to take place during one of the augmented periods of carriage that results from the imputed acquisition and clearance times. Figure 11.2 from Cauchemez et al (2006) illustrates the data augmentation approach. The model assumes conditional independence between consecutive observations of the same individual as well as between observations of different family members. The transmission model allows dependence of the process within the family.

**Table 11.6.** The number of observed changes in the individual carriage status over the observation intervals (Auranen et al 2000).

| Carriage among other family members at start of observation interval | Carriage | Age class 0–2 years: Carriage at next observation No | Age class 0–2 years: Carriage at next observation Yes | Age class 0–2 years Total | Age class 2–5 years: Carriage at next observation No | Age class 2–5 years: Carriage at next observation Yes | Age class 2–5 years Total |
|---|---|---|---|---|---|---|---|
| No carriage in in family | No | 562 | 33 | 595 | 107 | 10 | 117 |
| | Yes | 16 | 12 | 28 | 8 | 4 | 12 |
| Total | | 578 | 45 | 623 | 115 | 14 | 129 |
| At least one carrier in family | No | 24 | 1 | 25 | 12 | 6 | 20 |
| | Yes | 6 | 10 | 16 | 6 | 7 | 13 |
| Total | | 30 | 11 | 41 | 20 | 13 | 33 |

The transmission model is of the form in model (11.14). Let $\alpha$ be the rate to acquire carriage of serotype $s$ from community, $\beta$ be the rate of transmission from an infective in a family to susceptible, and $\mu$ be the clearance rate (no serotype specific rates), and $n$ be the size of the family. $C_i^{(s)}(t)$ is a 0,1 indicator if individual $i$ is a carrier of serotype $s$ at time $t$, $C_i(t)$ indicates any of the serotypes. Let $T_i$ denote the time of birth of individual $i$ , so that $t - T_i$ is the age of subject $i$. The intensities of the point processes of acquiring and clearing carriage are

$$\begin{aligned} \tilde{\lambda}_i^{(s)} &= \left[\alpha(t - T_i) + \beta(t - T_i) \sum_{k=1}^{n} C_k^{(s)}(t)\right] \\ &\quad \times \{1 - C_i(t)\} \\ \tilde{\mu}_i(t) &= \mu C_i(t) \end{aligned} \tag{11.15}$$

Model (11.15) makes several assumptions. The acquisition rates $\alpha$ and $\beta$ are assumed to depend on the age of the non-carrying susceptible. The acquisition rates are assumed to be the same for all three serotypes. The model allows carriage of only one serotype at a time. The duration of carriage is assumed to be an exponential random variable that is the same for all three serotypes. The model also requires an initial carriage state for each individual that is related to the proportion $\pi$ of Pnc carriers, assumed to be a single parameter across all age groups and for all serotypes. The proportion $\pi$ needed to be estimated because the initial observation was missing on some of the participants. The rate $\mu$ of clearing carriage was assumed to be $\mu_1$ for children under 2 years old and $\mu_2$ for family members over 2 years and older.

The full Bayesian model, notation, and methods for computation are presented in Auranen et al (2000). The parameters to be estimated were $\alpha$, $\beta$, $\mu$, and $\pi$ from the imputed acquisition and clearance times, the serotype data,

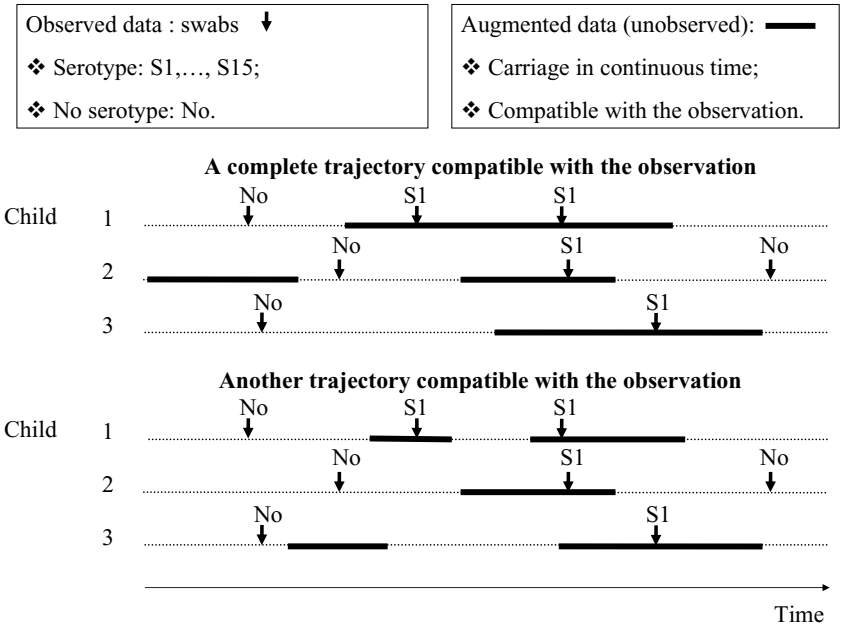


**Fig. 11.2.** Data augmentation strategy to estimate transmission parameters of *Streptococcus pneumoniae* in a longitudinal study of pneumococcal carriage. The observed data are presence or absence of pneumococcal bacteria in the nasopharynx. The data are augmented with the times of acquisition and clearance of carriage which give the period of carriage. The figure shows two different possible augmentations compatible with the observed data (from Cauchemez et al 2006d).

and the initial carriage status data. The priors on the parameters were assumed to be independent. Vague priors were used on all of the parameters, with means informed in part by the data. The acquisition rates $\alpha_f$ and $\beta_f$ were assumed to be same for all family members older than 5 years, with rate ratio $\varphi = \beta/\alpha$. An age dependent $\alpha(a)$ was assumed for children less than five years old. The function was formulated as piecewise constant over the interval 2 and 60 months of age. The rate ratio $\varphi = \beta(a)/\alpha(a)$ was assumed to be constant.

The age-dependent community acquisition rate in children under 5 years old increases up to approximately 0.3 new infections per year at age 18 months, which corresponds to the increased prevalence of carriage. Further estimates are in Table 11.7. The posterior mean of the ratio $\varphi$ of the within family (with at least one carrier in the family) and the community rate of acquisition was 25 in children under 5 years. In family members older than 5 years, the rate of community acquisition was 0.04 per year and the the rate ratio $\varphi_f$ was 15. The mean duration $(1/\mu_1)$ of carriage was longer in children less than 2

**Table 11.7.** Summary of the marginal posterior distribution of the parameters estimated for the Finnish pneumococcol carriage study in households. Parameter definitions are in the text (Auranen et al 2000).

| Parameter | Mean | Median | 90% equal-tail Credible interval |
|---|---|---|---|
| $\alpha_f$ (per year) | 0.037 | 0.037 | 0.016–0.061 |
| $\varphi$ | 25 | 23 | 14–44 |
| $\varphi_f$ | 15 | 10 | 3-42 |
| $\mu_1$ (per month) | 0.45 | 0.44 | 0.30–0.66 |
| $\mu_2$ (per month) | 0.71 | 0.69 | 0.49–1.01 |
| $1/\mu_1$ (months) | 2.3 | 2.3 | 1.5–3.3 |
| $1/\mu_2$ (months) | 1.5 | 1.4 | 1.0–2.0 |
| $\pi$ | 0.023 | 0.022 | 0.011–0.037 |

years old than in older family members ($1/\mu_2$). The model was assess using a number of approaches. A clear pattern in the data was the temporal clustering of pneumococcal carriage within families.

Cauchemez, et al (2006d) used a similar approach to Auranen et al (2000) to analyze a longitudinal study in France in schools rather than households described in Chapter 10.5.2. The primary scientific question of the analysis was whether the seven Pnc serotypes have a competitive advantage over the nonvaccine serotypes. The analysis proposed to study this question by comparing the mean duration of carriage ($1/\mu$) and within school child-to-child transmission rate ($\beta$) of the seven vaccine serotypes with those of the nonvaccine serotypes, denoted by V and U, respectively. No children had yet been vaccinated. The vaccine serotypes are 6B, 9V, 14, 18C, 19F, 23F. Vaccine serotype 4 was not included in the analysis because it was isolated in only 10 samples. The nonvaccine serotypes included in the analysis are 6A, 3, 19A, 11A, 15A, 23A, 17F, 10A, and 9L.

The model is similar to that in (11.15), except the term for transmission within school included the term $1/n$, where $n$ is the number of children in the school regardless of whether they took part in the study. The term $1/n$ serves as a density factor and reduces transmission the larger the school is. Thus the individual rate to acquire serotype $s$ at time $t$ was $\beta C_s(t)/n$, rather than just $\beta C_s(t)$, where $C_s(t)$ is the number of children carrying serotype $s$ at time $t$.

An expression for the secondary attack rate, SAR, the probability that a child carrying the bacteria transmits to a non-carrying child during the carriage period, is presented. Comparing the SAR for the vaccine types to that of the nonvaccine types allows comparison of the overall fitness for transmission since it is a function of the mean duration of carriage and the child to child transmission rate. Let $L$ be the duration of carriage with density $f(L) = \mu \exp(-L\mu)$ and $1 - \exp(-\beta L/n)$ be the probability that a carrying child during time period $L$ transmits to a non-carrying child in the school.

**Table 11.8.** Summary of the marginal posterior distribution of the parameters estimated for the study of pneumococcal carriage in schools in France. Parameter definitions are in the text (Cauchemez et al 2006).

| | Vaccine serotype | | Non-vaccine serotype | | Ratio | |
|---|---|---|---|---|---|---|
| | Mean | 95% credible interval | Mean | 95% credible interval | Mean | 95% credible interval |
| $1/\mu$ (days) | 23 | 21–25 | 22 | 20–24 | 1.06 | 0.94–1.28 |
| $\beta$ (% day$^{-1}$) | 4.6 | 4.2–5.0 | 5.1 | 4.5–5.6 | 0.91 | 0.80–1.05 |
| SAR (%) | | | | | | |
| $n = 30$ | 3.4 | 3.2–3.7 | 3.6 | 3.3–3.8 | 0.97 | 0.88–1.06 |
| $n = 50$ | 2.1 | 1.9–2.2 | 2.2 | 2.0–2.3 | 0.97 | 0.88–1.06 |
| $n = 100$ | 1.1 | 1.0-1.2 | 1.1 | 1.0–1.2 | 0.97 | 0.88–1.06 |

Then

$$SAR = \int_0^\infty (1 - \exp(-\beta L/n) f(L) dL,$$

which reduces to $SAR = (1+N\mu\beta^{-1})^{-1}$. The results are summarized in Table 11.8. There was no evidence that the vaccine serotypes had different transmission characteristics than the nonvaccine serotypes. (Note: Cauchemez et al (2006) define $\mu$ as the mean duration of carriage, not $1/\mu$). Cauchemez et al (2006c) further investigated heterogeneity among the 15 serotypes using a clustering step to select a parsimonious description of the transmission characteristics.

### 11.4.2 Markov transition model

Another approach to estimating the acquisition parameters $\alpha$ and $\beta$ in equation (11.14) and the clearance rate $\mu$ is by explicit formulation of the transition matrix for the Markov process in the households (Auranen et al 1996). The spread of carriage within the family is modeled by the transition between states in a Markov process. The state of the family at time $t$ is the combined state of the individuals in the family. The number of possible states depends on the number of individuals in the family. In a family with three members, there are eight possible states. Letting 1 indicate a carrier and 0 a non-carrier, the possible states are 000, 001, 010, 011, 100, 101, 110, and 111. A family in which at one observation time, the second person is a carrier, the other two non-carriers, and at the next observation time, both the second and third are carriers, makes the transition from state 010 to state 011. The corresponding intensity matrix $Q$ with elements $q_{ij}$ is

| | 000 | 001 | 010 | 011 | 100 | 101 | 110 | 111 |
|---|---|---|---|---|---|---|---|---|
| 000 | $q_{11}$ | $\alpha^{(3)}$ | $\alpha^{(2)}$ | 0 | $\alpha^{(1)}$ | 0 | 0 | 0 |
| 001 | $\mu^{(3)}$ | $q_{22}$ | 0 | $\alpha^{(2)}+\beta^{(2)}$ | 0 | $\alpha^{(1)}+\beta^{(1)}$ | 0 | 0 |
| 010 | $\mu^{(2)}$ | 0 | $q_{33}$ | $\alpha^{(3)}+\beta^{(3)}$ | 0 | 0 | $\alpha^{(1)}+\beta^{(1)}$ | 0 |
| 011 | 0 | $\mu^{(2)}$ | $\mu^{(3)}$ | $q_{44}$ | 0 | 0 | 0 | $\alpha^{(1)}+2\beta^{(1)}$ |
| 100 | $\mu^{(1)}$ | 0 | | | $q_{55}$ | $\alpha^{(3)}+2\beta^{(3)}$ | $\alpha^{(2)}+\beta^{(2)}$ | 0 |
| 101 | 0 | $\mu^{(1)}$ | 0 | 0 | $\mu^{(3)}$ | $q_{66}$ | 0 | $\alpha^{(3)}+2\beta^{(2)}$ |
| 110 | 0 | 0 | $\mu^{(1)}$ | 0 | $\mu^{(2)}$ | 0 | $q_{77}$ | $\alpha^{(3)}+\beta^{(3)}$ |
| 111 | 0 | 0 | 0 | $\mu^{(1)}$ | 0 | $\mu^{(2)}$ | $\mu^{(3)}$ | $q_{88}$ |

The elements on the diagonals represent the intensity of staying in the same state. The $q_{ii} = -\sum_{j\neq i} q_{ij}$ (Karlin and Taylor 1975). The element (4,8) represents the transition from state 011 to state 111. Individual number 1 has the transition rate $\alpha^{(1)} + 2\beta^{(1)}$, representing the rate of acquisition from the community and the exposure by the two carriers in the household. The model only allows one family member to make a transition during the time period of the time step. The 0s in the matrix represent transitions that are not allowed, thus have 0 intensity. The matrix can represent households of any size and include age- or covariate- (vaccination status) dependent rates. For the Hib analysis, Auranen et al (1996) had two levels of the three rate parameters, one for children under 7 years old and one for everyone older than 7. They used a Bayesian approach to estimating the parameters.

Melegaro, et al (2004) used a similar model to estimate the acquisition and clearance rates for a household study of pneumococcal carriage in the United Kingdom (Hussain et al 2005) presented in Chapter 10.5.3. A density correction factor $(z-1)^w$ was added, where $(n-1)$ is the number of other family members in a household size $n$, and $w$ corresponds to the level of density dependence. When $w = 0$, the model represents density-independent transmission.

The population was divided in children under 5 years and everyone else greater than 5 years, denoted by $i = 1, 2$. $C_1(t)$ and $C_2(t)$ are number of carrier children ($< 5$ yrs) and adults in household. Then the probability of a transition in a short time interval $\delta t$ for an individual in age class $i$ is

$$\Pr_i(S \to C)_{\delta t} = \left(\alpha_i + \frac{\beta_{1i}C_1(t) + \beta_{2i}C_2(t)}{(n-1)^w}\right) \cdot \delta t,$$

$$\Pr_i(C \to S)_{\delta t} = \mu_i \cdot \delta t.$$

They use a Markov model with 1-day intervals to analyze 28-day interval data assuming only one person can change in household per day. The parameters were estimated using a likelihood approach. The estimate of the density

parameter $w$ was significantly greater than 0 ($w = 1.2$, 95% CI 0.2–2.2) suggesting that transmission within households might depend on density.

In a further analysis, Melegaro et al (2007) extended the model to estimate serotype-specific transmission parameters. Five distinct data sets were constructed, one for each of the target serotypes. The carriage status of each study participant was recorded at each monthly visit as 0 (noncarrier), 1 (carrier of the target serotype), 2 (carrier of any other serotype), or 9 when the swab was not taken or the laboratory result was not reported. Estimation used a likelihood approach.

## Problems

**11.1. $\mathbf{VE}_{S,p}$ and $\mathbf{VE}_{S,SAR}$**
(a) The first part of the problem is described here.
(b) The second part of the problem is described here.

**11.2. Comparing antibody efficacies**
(a) Compute the antibody efficacies based on the SAR, CPI, and AR for children and adults using equation 11.9 and the results in Tables 10.5 and 11.2.
(b) Compare the estimates and explain why the antibody efficacy based on the SAR is much higher than those based on the CPI and AR.

**11.3. Problem Heading**
(a) The first part of the problem is described here.
(b) The second part of the problem is described here.

# 12

# Analysis of Independent Households

In this chapter, we consider methods of analysis that assume that the households are independent of one another. The most commonly used approach is to estimate vaccine efficacy based on the conventional secondary attack rate. We also consider the estimation of indirect effects of vaccination using household studies.

## 12.1 Conventional SAR Analysis

### 12.1.1 Setting up the secondary attack rate analysis.

To estimate the conventional household SAR, the main task is to set up the analysis. Decisions need to be made on

1. who in the household is eligible to be a secondary case, and
2. who of the eligible household members are the secondary cases.

The first contributes to the denominator of the secondary attack rate, the second contributes to the numerator.

Data on the time of onset of disease for each case in the household as well as knowledge of who is susceptible are required. Then one decides who is the index, or primary case in the household, and which of the other cases in the household could reasonably have been infected by the primary case. Occasionally one differentiates the index case, the case that results in ascertainment of the household, from the primary case, the earliest temporal case in the household. To decide which of the subsequent cases in the household could have been infected by the primary case, one needs estimates or assumptions about the minimum and maximum incubation periods, the latent period and its relation to the incubation period, and the maximum time that a person remains infectious. These values will vary according to the disease of interest. Using this information, one then needs to define the time interval after the primary case that would include the secondary cases. Based on the time of

onset data within each household, each case is defined as being either a secondary case or not. Co-primary cases are people who developed disease too soon after the primary case to have been infected by the primary case. They are not counted as secondary cases, and are generally simply excluded from the analysis. They are not included in the denominator of the SAR. Primary cases are also not included in the analysis. The estimated household secondary attack rate is the total number of secondary cases in all households divided by the total number of at-risk susceptibles in all households. In some cases, tertiary or higher generation cases may be included in the analysis by calling the secondary cases the primary case for further chains of transmission. Tertiary cases or higher cases are included in the denominator of exposed individuals for the secondary attack rate, but not in the number of cases in the numerator. Chapter 10 contains several examples of studies with intervals for determining the eligible susceptible household members and the secondary cases.

**Pertussis vaccination** Préziosi and Halloran (2003) defined exposed susceptibles as children with no history of pertussis living in a compound with an index case (Section 10.2.3) . Onset of pertussis symptoms was assumed to be the onset of infectiousness, thus the latent period was assumed to equal the incubation period. Co-primaries were those cases whose onset of cough was $<$7 days after that of the index case. To allow for uncertainty in duration of infectiousness, a secondary case was defined as a case whose date of onset was $\geq$7 days after that of the index case and less than a variable cutoff, specifically none, 56, 42, or 28 days. Similar assumptions were made by Kendrick and Eldering (1939) (Section 10.2.2). In the PHLS Epidemiologic Research Laboratory (1982), the co-primaries were those within seven days of the index case and secondary cases were those that occurred within about 42 days of the index case and at least seven days after the index case (Section 10.2.4). In the re-analysis of this study, Fine, et al (1988), co-primaries were defined as cases within one week of the primary case. Incidence cases were those that occurred more than one week after the primary cases. These included more than potentially secondary cases.

**Measles** For measles, Orenstein et al (1985) recommend 18 days of follow-up in a household after the onset of rash in the primary case. Garenne et al (1993) defined secondary cases as those occurring in the same compound 7 to 18 days after the index case (Section 10.4.1). Exposed susceptibles were children who had never had measles living in a compound where there was a clinical case.

**Mumps** In a secondary attack rate analysis of mumps vaccine efficacy in an outbreak investigation, Kim-Farley et al (1985) defined co-primaries as cases in family members occurring within 10 days of the onset of disease in the index case. Cases with onset of disease more than 30 days after the index case were considered tertiary cases. Children with previous history of mumps disease, unknown vaccine histories, or unknown dates of vaccination were excluded from the analysis.

In the conventional secondary attack rate analysis, the assumption is that the households or other small transmission units are independent of one another. There is an asymmetric assumption that the index case and co-primaries get infected from outside the unit, while the other susceptibles are exposed only within the unit. This assumption is very different than the assumption in Chapter 11 in which individuals can acquire infection from the community even if there is an infective in the household. If the transmission probability or secondary attack rate is estimated without taking into account the opportunity to become infected outside of the transmission unit, it will overestimate the actual probability of becoming infected per contact. In general, ratio measures, such as the vaccine efficacy based on the ratio of secondary attack rates, are less biased by this problem. Kemper (1980) discusses biases in conventional SAR estimation. The drawback in using models such as those in Chapter 11 is that they contains strong modeling assumptions about the mixing in the community. An advantage of the conventional SAR studies or case-contact study designs is that they do not make assumptions about the community at large. They are also quite simple to compute once the biological assumptions about the time intervals containing the secondary cases have been made.

### 12.1.2 Vaccine efficacy from conventional SAR

As described in Chapter 2, the secondary attack rates can be differentiated by the vaccine status of the primary case and/or the vaccine status of the secondary cases. In general, there are at least seven measures potentially of interest. Considering the estimates of VE based on the relative secondary attack rates, there are three main unstratified vaccine effects:

$$VE_{S.1/.0} = 1 - \frac{SAR_{.1}}{SAR_{.0}}\ , \quad VE_{I1./0.} = 1 - \frac{SAR_{1.}}{SAR_{0.}}\ ,$$
$$VE_T = 1 - \frac{SAR_{11}}{SAR_{00}}\ . \tag{12.1}$$

The traditional measure of $\mathrm{VE}_{S,SAR}$ to estimate the protective effect of vaccination in household studies corresponds to $\mathrm{VE}_{S.1/.0}$. When not ambiguous, we use the notation $\mathrm{VE}_{S,SAR}$ for traditional estimates. If one stratifies on the vaccine status of the infective person or the susceptible person, then there are four further stratified measures of $\mathrm{VE}_S$ and $\mathrm{VE}_I$:

$$VE_{S01/00} = 1 - \frac{SAR_{01}}{SAR_{00}}, \quad VE_{S11/10} = 1 - \frac{SAR_{11}}{SAR_{10}},$$
$$VE_{I10/00} = 1 - \frac{SAR_{10}}{SAR_{00}}, \quad VE_{I11/01} = 1 - \frac{SAR_{11}}{SAR_{01}}. \tag{12.2}$$

The confidence interval for any of the $\mathrm{VE}_{SAR}$ measures is generally based on log relative risk. If $a_0$, $a_1$ are the number of cases and $n_0$, $n_1$ are the

number exposed in the relevant comparison groups, then the relative risk is $RR = a_1 n_0 / a_0 n_1$. The standard deviation of the log relative risk is

$$\sigma = \left( \frac{1}{a_1} - \frac{1}{n_1} + \frac{1}{a_0} - \frac{1}{n_0} \right)^{\frac{1}{2}}.$$

The vaccine efficacy estimate and 95% confidence interval are

$$VE_{SAR} = 1 - \frac{a_1/n_1}{a_0/n_0}, \tag{12.3}$$

$$95\%\ \text{CI} \quad [1 - \exp(\log(RR) + 1.96 * \sigma), 1 - \exp(\log(RR) - 1.96 * \sigma)].$$

For example, in the Medical Research Council study of pertussis vaccination (Table 10.2), there were $a_1 = 37$ pertussis cases among $n_1 = 203$ home exposures in the vaccinated children, and $a_0 = 151$ cases among $n_0 = 173$ home exposures in the unvaccinated children. In the report, the vaccination status of the exposing children is not included, so only the traditional, unstratified $\text{VE}_{S,SAR}$ can be computed. The $\text{VE}_{S,SAR}$ estimate and 95% confidence interval are 0.79 [95% CI 0.72, 0.84].

This simple approach does not take into account that several children may be exposed to the same infective, so that there may be correlation within households. Quite often, one infectious person exposes several people, possibly within a transmission unit, such as a household. Correlation within transmission units or unmeasured heterogeneity across transmission units could result, for example, from differences in infectivity, difference in mixing within the unit, or genetic variation. This conventional method to estimate the confidence intervals for vaccine efficacy fails to take the structure of the clustered binary data into account.

## 12.2 SAR analysis taking correlation into account

Preźiosi and Halloran (2003) and Halloran, Préziosi, and Chu (2003) were particularly interested in estimating the effect of pertussis vaccination on reducing infectiousness of vaccinated cases, $\text{VE}_I$. They analyzed the pertussis vaccination study in Niakhar, Senegal, described in Chapter 10.2.3. The data are summarized in Table 10.1. Many of the compounds had several children, so correlation within compound might be important. They developed methods for estimating the VE measures based on the SAR that take correlation within transmission units into account. We present these methods using the pertussis vaccination study in Niakhar, Senegal, as an example.

### Notation

Let $n$ be the number of compounds with a unique index case and $m_i$ be the number of susceptibles in the $i$th compound. Let $y_{ij}$ be the binary (0,1)

pertussis outcome of the $j$th susceptible exposed to the index case in the $i$th compound for any given case definition. Let $\boldsymbol{x}_{ij} = (x_{ij1}, \cdots, x_{ijp})'$ denote a $p \times 1$ vector of explanatory variables associated with $y_{ij}$. In particular, let $x_{i\cdot 1}$ denote the vaccine status of the index case in compound $i$, and $x_{ij2}$ the vaccine status of the $j$th exposed susceptible individual in compound $i$. Complete pertussis vaccination requires at least three doses of vaccine. This analysis considers only unvaccinated and fully vaccinated children, with $x_{i\cdot 1} = 0$ and $x_{i\cdot 1} = 1$ for an unvaccinated and fully vaccinated index case. Similarly, $x_{ij2}$ is 0 or 1 for the unvaccinated and fully vaccinated susceptibles.

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

$$
\begin{aligned}
a_{vs} &= \sum_{i=1}^{n}\sum_{j=1}^{m_i} I_{V=v} I_{S=s} y_{ij}, \quad N_{vs} = \sum_{i=1}^{n}\sum_{j=1}^{m_i} I_{V=v} I_{S=s} \ , \\
a_{\cdot\cdot} &= \sum_{i=1}^{n}\sum_{j=1}^{m_i} y_{ij}, \quad N_{\cdot\cdot} = \sum_{i=1}^{n} m_i \ , \\
a_{\cdot s} &= \sum_{i=1}^{n}\sum_{j=1}^{m_i} I_{S=s} y_{ij}, \quad N_{\cdot s} = \sum_{i=1}^{n}\sum_{j=1}^{m_i} I_{S=s} \ , \\
a_{v\cdot} &= \sum_{i=1}^{n}\sum_{j=1}^{m_i} I_{V=v} y_{ij}, \quad N_{v\cdot} = \sum_{i=1}^{n}\sum_{j=1}^{m_i} I_{V=v} \ .
\end{aligned}
$$

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

The three main unstratified nonparametric VEs in equation (12.1) and stratified nonparametric VEs in equation (12.2) can be estimated using these SARs with the standard confidence intervals as in equation(12.3).

### 12.2.1 Vaccine efficacy based on the logistic model

To take correlation within compounds into account, a marginal model or a random-effects model could be used. The parametric form in both cases is

the logistic model, with the SAR as the usual probability $p$ of an event. The model-based approach allows inclusion of covariates, such as age, either of the index case as compound-level environmental variables or of the susceptibles as individual variables. Diggle, Liang and Zeger (1994, pp. 131-135) explain that in marginal models, inference about population averages is the focus. If there is heterogeneity across compounds in the baseline transmission, then the estimated baseline coefficients represent an average over the heterogeneities. The correlation structure is some function of the marginal mean and possibly additional parameters.

In the random-effects model, a slightly different baseline transmission is estimated for each compound, with the degree of heterogeneity estimated in the variance of the random effect. The vaccine effects in each compound are interpreted in relation to that compound's baseline transmission. In considering vaccine efficacy, out primary scientific question is about the population average, or marginal, VE measures. Thus, the marginal model is the model of choice. The coefficients for the marginal and random-effects models are indicated by $\boldsymbol{\beta}$ and $\boldsymbol{\beta}^*$

### The marginal model

The marginal model for the logit of the $\text{SAR}_{ij}$ of the $j$th person in the $i$th household is

$$\text{logit}(SAR_{ij}) = \beta_0 + \beta_1 x_{i\cdot 1} + \beta_2 x_{ij2} \ , \tag{12.4}$$

where $x_{i\cdot 1}$ denotes the vaccine status of the index case in compound $i$ and $x_{ij2}$ denotes the vaccine status of the $j$th exposed susceptible in compound $i$. The vaccine status of the index case, $x_{i\cdot 1}$, enters the analysis as a compound-level, environmental variable. Because we are interested in VE estimates on the SAR scale, we transform the parameters from the logistic model to the probability scale. The stratified SARs from model (12.4) are

$$\begin{aligned} SAR_{00} &= \frac{\exp\beta_0}{1+\exp\beta_0}, \quad SAR_{03} = \frac{\exp\left(\beta_0+\beta_2\right)}{1+\exp\left(\beta_0+\beta_2\right)}, \\ SAR_{30} &= \frac{\exp\left(\beta_0+\beta_1\right)}{1+\exp\left(\beta_0+\beta_1\right)}, \quad SAR_{33} = \frac{\exp\left(\beta_0+\beta_1+\beta_2\right)}{1+\exp\left(\beta_0+\beta_1+\beta_2\right)}. \end{aligned} \tag{12.5}$$

Parameter estimates from the above model provide estimates for the stratified $\text{VE}_{S00/03}$ and $\text{VE}_{S30/33}$, the stratified $\text{VE}_{I00/30}$ and $\text{VE}_{I03/33}$, as well as $\text{VE}_T$. Plugging the expressions for the SARs into equations (12.2), and for $\text{VE}_T$ in equation (12.1), the expressions for the VE measures are

$$VE_{S03/00} = \frac{1-\exp(\beta_2)}{1+\exp(\beta_0+\beta_2)}, \quad VE_{S33/30} = \frac{1-\exp(\beta_2)}{1+\exp(\beta_0+\beta_1+\beta_2)},$$
$$VE_{I30/00} = \frac{1-\exp(\beta_1)}{1+\exp(\beta_0+\beta_1)}, \quad VE_{I33/03} = \frac{1-\exp(\beta_1)}{1+\exp(\beta_0+\beta_1+\beta_2)},$$
$$VE_T = \frac{1-\exp(\beta_1+\beta_2)}{1+\exp(\beta_0+\beta_1+\beta_2)}. \tag{12.6}$$

To obtain estimates of the unstratified $\mathrm{VE}_{I3./0.}$ and $\mathrm{VE}_{S.3/.0}$, we fit additional submodels, such as $\mathrm{logit}(SAR_{ij}) = \beta_0' + \beta_1' x_{i\cdot 1}$ and $\mathrm{logit}(SAR_{ij}) = \beta_0'' + \beta_2'' x_{ij2}$ and transform back to get

$$VE_{I3./0.} = \frac{1-\exp(\beta_1')}{1+\exp(\beta_0'+\beta_1')}, \quad VE_{S.3/.0} = \frac{1-\exp(\beta_2'')}{1+\exp(\beta_0''+\beta_2'')}. \tag{12.7}$$

Alternatively, one could use the parameter estimates from the full model (12.4) and substitute the respective means of $x_{i\cdot 1}$ and $x_{ij2}$. We estimated the marginal model taking correlation of transmission within compound into account using generalized estimating equations (GEE) (Liang and Zeger 1986). The analysis was done using the `repeated` option in PROC GENMOD in SAS version 8.2 (SAS Institute I, 1999) assuming an exchangeable working correlation matrix.

Appropriate confidence intervals on the transformed scale are obtained using the bootstrap (Efron and Tibshirani 1993). Bootstrap samples were selected using the compound as the sampling unit. The GEE logistic regression coefficients were estimated for each bootstrap sample, then transformed to the probability scale to get the VE estimates for that bootstrap sample. Three different bootstrap confidence intervals were computed, namely the percentile, the bias-corrected (BC), and the bias-corrected and accelerated ($\mathrm{BC}_a$) intervals. Confidence intervals were based on 2000 bootstrap samples (Efron and Tibshirani 1993, p. 275). Bootstrap confidence intervals sampling on compounds were also computed for the VE estimators based on the nonparametric SARs described in the previous section. Analytic confidence intervals for the GEE estimates of VE on the transformed scale were obtained using the multivariate delta method (Agresti 1990).

### The random-effects model

The random effects model for the logit of the $\mathrm{SAR}_{ij}$ of the $j$th person in the $i$th household is

$$\mathrm{logit}(SAR_{ij}|U_i) = (\beta_0^* + U_i) + \beta_1^* x_{i\cdot 1} + \beta_2^* x_{ij2}\ . \tag{12.8}$$

The simplest model assumes that the random effect $U_i \sim N(0, \sigma^2)$. On the logistic scale, the parameter $\beta_0^*$ would be interpreted as the log odds of transmission from an unvaccinated index case to an unvaccinated susceptible for a typical compound with random effect $U_i = 0$. The parameter $\beta_1^*$ would be

the log odds ratio of transmission occurring when the index case is vaccinated compared to when it is unvaccinated within any given compound. The parameter $\beta_2^*$ would be the log odds ratio of transmission occurring when a susceptible in the compound is vaccinated compared to a susceptible in that same compound who is unvaccinated.

However, our interest is in transforming to the SAR scale to obtain the different VE estimates. The compound-specific $\text{SAR}_{ij}$s are obtained by incorporating the random effect into the expression. For example, the compound-specific $\text{SAR}_{00i}$ from an unvaccinated index case to an unvaccinated susceptible is $SAR_{00i}|U_i = \exp(\beta_0^* + U_i)/[1 + \exp(\beta_0^* + U_i)]$. The marginal $\text{SAR}_{00}$ is the estimated expectation of the $\text{SAR}_{00}$ obtained by numerical integration over the estimated distribution of the random effects. The $\text{VE}_i$ estimates for each compound are obtained from expressions analogous to (12.6). The marginal VE estimates are the estimated expectations obtained by numerical integration over the estimated distribution of the random effects. To obtain estimates of the unstratified $\text{VE}_{I3./0.}$ and $\text{VE}_{S.3/.0}$, we fit random effects submodels similar to those described above.

Two methods were used to estimate the random effects model. The first is a Bayesian hierarchical model(Carlin and Louis 2000), the second is a nonlinear mixed model (Davidian and Giltinan 1995). Computation was done using Markov chain Monte Carlo (MCMC) methods in `WinBUGS` (Spiegelhalter, Thomas, Best 2000). The population mean VE measures were computed by averaging over the compounds at each iteration. The 95% posterior credible intervals for the VE measures are available directly on the transformed scale from the approximation to the posterior distribution from the MCMC chains. The nonlinear mixed model was fit using PROC NLMIXED (Wolfinger 1999) in SAS verion 8.2 (SAS Institute I, 1999). Details are in Halloran et al (2003).

### 12.2.2 Pertussis vacccine efficacy

Estimates of the baseline $\text{SAR}_{00}$ for each model are shown in Figure 12.1. Horizontal lines represent point estimates of the nonparametric pooled $\text{SAR}_{00}$, the marginal GEE model, and the estimated expected $\text{SAR}_{00}$ on the transformed scale for the two random effects models. Included for comparison are horizontal lines for the estimate of the nonlinear mixed model and the mean from the Bayesian MCMC chain of the $\text{SAR}_{00i}$ when $U_i = 0$.

For the random effects models, the $\text{SAR}_{00i}$s for each of the 109 compounds are shown in Figure 12.1. For the nonlinear mixed model, the point estimates of the $\text{SAR}_{00i}$ are shown. For the Bayesian model, the mean $\text{SAR}_{00i}$ and 95% posterior CI from the MCMC chain are shown. The individual estimated $\text{SAR}_{00i}$s range from about 0.1 to 0.7, though 51 of the 109 compounds had no secondary cases.

Figure 12.2 shows the point estimates and histograms of 2000 bootstrap estimates of the $\text{VE}_I$, $\text{VE}_S$, and $\text{VE}_T$ parameters based on the GEE model. Figure 12.3 shows the different point estimates and confidence intervals for

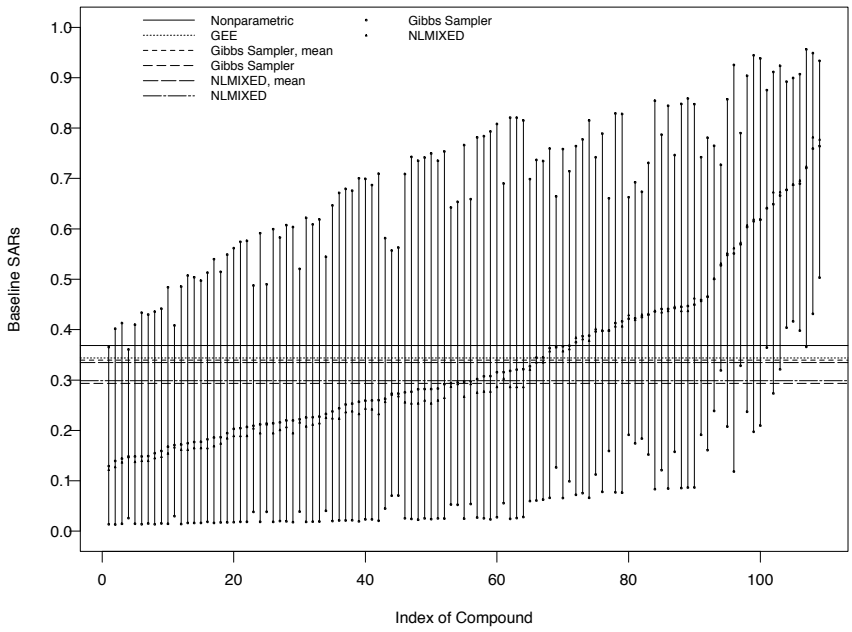


**Fig. 12.1.** Baseline $SAR_{00}$ for each model represented by horizontal lines. Compound-level baseline $SAR_{00i}$'s for the random-effects nonlinear mixed model (NLMIXED) are also plotted. The posterior mean and 95% posterior CI are plotted for each Bayesian compound-level $SAR_{00i}$. The indexes of the compounds are in order of increading mean from the MCMC chain. ——- nonparametric; · · · GEE; – – – – Gibbs sampler, mean; — — — Gibbs sampler; —- —- —— NLMIXED, mean; —- - —- - NLMIXED; · Gibbs sampler; · NLMIXED. (from Halloran et al 2003)

$VE_S$, $VE_I$, and $VE_T$. Table 12.1 contains selected results. The point estimates for $VE_I$ and $VE_T$ obtained from the nonparametric SAR and from the GEE are nearly identical. The bootstrap CIs for the nonparametric VE estimates are wider than the simple CIs based on the log relative risk. In particular, the bootstrap CIs for $VE_I$, and to a lesser extent, $VE_T$ are wider. For example, the BC bootstrap 95% CI of $VE_{I3./0.}$ is 1.94 wider than the simple 95% CI. The difference is less pronounced with CIs of $VE_S$, with the ratio of the lengths being between 1.2 and 1.3. Thus, the conventionally used CI substantially underrepresents the variability in the data. The greater sensitivity of the variability of the $VE_I$ and $VE_T$ estimators to compound-level effects might result from the vaccine status of the index case being a compound-level environmental variable. The nonparametric estimate of $VE_{S33/30}$ is unstable because the total number of secondary cases was only 20, compared with 134

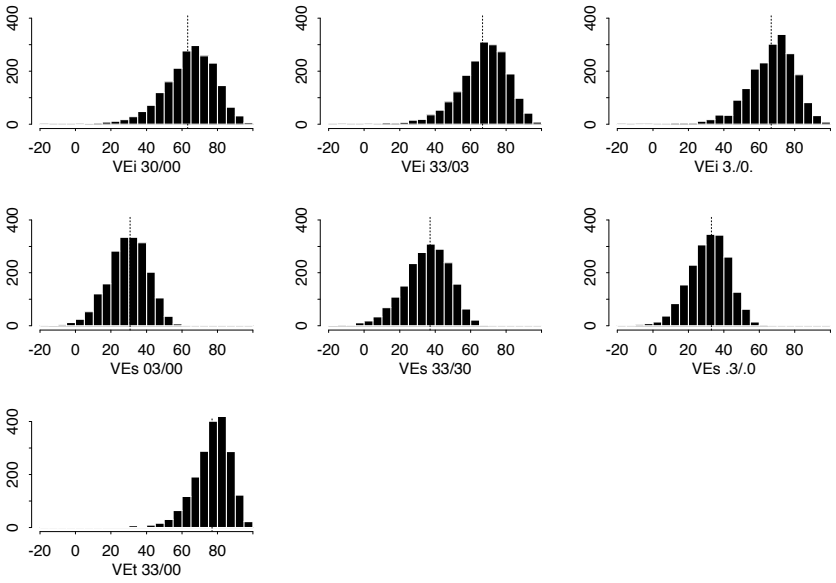


**Fig. 12.2.** Histograms of 2,000 bootstrap estimates of (top row) VE for infectiousness, $\text{VE}_I$ stratified and unstratified; (middle row) VE for susceptibility, $\text{VE}_S$, stratified and unstratified; and (bottom) total VE, $\text{VE}_T$, based on the GEE logistic regression parameters. The dotted line in each histogram indicates the estimate for the actual dataset (from Halloran et al 2003).

cases for $\text{VE}_{S03/00}$, so both the simple and the BC bootstrap CIs are quite wide.

The bootstrap CIs of the GEE estimates of $\text{VE}_I$ are also wider than those based on the simple CI for the nonparametric VE estimates, however, not as much wider as the bootstrap CIs of the nonparametric VE estimates. For example, the GEE percentile, BC, and $\text{BC}_a$ bootstrap 95% CIs for $\text{VE}_{I3./0.}$ compared to the simple SAR 95% CI are 1.63, 1.74, and 1.83 wider, respectively. Thus, the parametric model in the GEE helps stabilize the estimation compared to the nonparametric approach.

The multivariate delta method CIs on the GEE estimates are symmetric and similar in length to the percentile bootstrap CIs. However, the normality assumption of the $\text{VE}_I$ and $\text{VE}_T$ estimators is clearly violated, so we do not recommend using the multivariate delta method. Also, CIs based on the multivariate delta method could theoretically exceed one, which could cause difficulty since vaccine efficacy is bounded at 1.

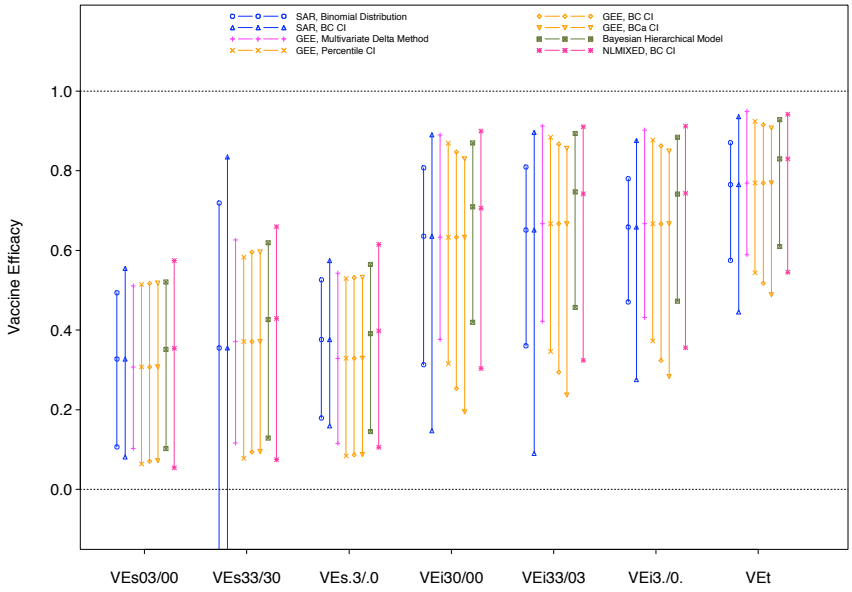


**Fig. 12.3.** Comparison of approaches to estimating SARs and confidence intervals. Halloran et al (2003)

**Table 12.1.** Pertussis vaccine efficacy estimates from the Niakhar region, Senegal, 1993. (from Halloran, Préziosi and Chu 2003)

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

* BC = bias-corrected bootstrap confidence interval

**Table 12.2.** Antiviral Efficacies for Oseltamivir (Halloran et al 2007)

| | based on laboratory-confirmed infection with symptoms | | | |
|---|---|---|---|---|
| Effect | AVE.$_d$ | 95% C.I. | Drug | Control |
| AVE$_S$ = 1−SAR$_{11}$/SAR$_{10}$ (Osel I alone) | | | | |
| 1 − 7 Days | 81 | 35, 94 | 3/237 | 16/241 |
| 2 − 7 Days | 81 | 35, 94 | 3/237 | 16/241 |
| AVE$_S$ = 1−SAR$_{01}$/SAR$_{00}$ (Osel II alone) | | | | |
| 1 − 7 Days | 91 | 64, 98 | 2/205 | 22/195 |
| 2 − 7 Days | 91 | 62, 98 | 2/205 | 21/194 |
| AVE$_I$ = 1−SAR$_{10}$/SAR$_{00}$ (Osel I/Osel II) | | | | |
| 1 − 7 Days | 81 | 45, 93 | 4/180 | 22/190 |
| 2 − 7 Days | 80 | 43, 93 | 4/180 | 21/189 |
| AVE$_T$ = 1−SAR$_{11}$/SAR$_{00}$ (Osel I/Osel II) | | | | |
| 1 − 7 Days | 91 | 63, 98 | 2/195 | 22/190 |
| 2 − 7 Days | 91 | 61, 98 | 2/195 | 21/189 |

### 12.2.3 Varying case definition and cutoff

Préziosi and Halloran (2003) considered the effect of varying the case definition and the cutoff date on the seven VE estimates. The primary focus of the analysis was on estimating VE$_I$. The primary method of analysis was the GEE approach using the bias-corrected and accelerated (BCa) bootstrap confidence intervals described in the previous section. Based on the main case definition and no cutoff of secondary cases, vaccine efficacy for infectiousness VE$_I$ was estimated to be 0.85 (95% CI 0.46–95) for children vaccinated with three doses of a whole cell (94%) or an acellular (6%) pertussis vaccine.

## 12.3 Estimating influenza antiviral efficacies

Halloran, et al (2006) used the conventional secondary attack rate to analyze the four randomized household studies of influenza antiviral efficacies presented in Chapter 10. Each of the efficacies can be based on 1) laboratory-confirmed influenza illness, AVE.$_d$, or 2) laboratory-confirmed infection, AVE.$_i$, in the eligible contacts. Here we present only the estimates based on laboratory confirmed influenza illness, AVE.$_d$, for the two oseltamivir studies (Table 12.2). As discuss in Chapter 10, the randomization schemes in the studies constrained which SARs, and thus which AVEs could be estimated. Influenza has a very short incubation period. The interval for co-primaries was assumed to be either 1 day or 2 days after ascertainment of the index case.

## 12.4 Mini-community Designs for Indirect Effects

In the minicommunity design, the household or other small transmission unit serves as the unit in which to estimate indirect effects of vaccination, similar to studies in larger communities to estimate indirect, total, and overall effects (Chapter 13). The gradient from small transmission units, such as households, to compounds as in Niakhar, to day care centers, to schools, to towns or whole countries is fairly continuous. Thus, this section could also have been contained in Chapter 13, but here we focus on households. Similar to many household-based vaccine efficacy studies, these minicommunity studies can be nested in either randomized clinical trials or observational studies where the primary analysis is based on unconditional measures. Unlike the other efficacy measures in this chapter, the estimates of the indirect effects of vaccination do not condition on the index case being a case of infection or disease. In the indirect effect measures, the analysis conditions only on the vaccination status of the index child or children in the household. The outcomes of interest are the disease or infection status of the other members of the household. Then the estimates of the indirect effects in the other members of the household are based on one of the unconditional risk measures, such as attack rate or case per person time in the other members of the household. If based on incidence rates per person-time, then

$$\mathrm{VE}_{IIA} = 1 - \frac{\dfrac{\text{no. of cases in household members of vaccinated children}}{\text{person-time of household members of vaccinated children}}}{\dfrac{\text{no. of cases in household members of unvaccinated children}}{\text{person-time of household members of unvaccinated children}}}. \tag{12.9}$$

### 12.4.1 Pertussis

Trollfors et al (1998) nested a study of the indirect effects of pertussis vaccination in households in the Swedish study described in (Section 10.2.5) They estimated the indirect effects based on equation (12.9) using the ratio of the incidence rates (pertussis cases divided by total time at risk) in parents and younger siblings of recipients of DTaP or DT. They used four different case definitions, the first being similar to the WHO definition, the second based on criteria developed by themselves. They further divided it by $\geq 21$ days of paroxysmal cough and cough $\geq 7$ days. Other criteria were similar to those discussed in Section 12.2.3. The results are in Table 12.3. Unfortunately, the original paper does not include the amount of person-time computed for each group.

Individually randomized studies in which a small transmission unit is accrued into the study, such as a family for a mini-community study, or an augmented study in which the partner is not randomized to vaccine or control are essentially the same randomization.

**Table 12.3.** Number of pertussis cases in parents and younger siblings of study children and indirect protection achieved by vaccination of the study child with pertussis toxoid (from Trollfors et al 1998)

| | Pertussis Cases | | Indirect protection (%) |
|---|---|---|---|
| | DTaP | DT | |
| Parents | | | |
| WHO definition | | | |
| $\geq$ 21 days of paroxysmal cough | 11 | 26 | 60 (16–82) |
| $\geq$ 7 days of cough | 23 | 35 | 38(−9–65) |
| Göteborg definition | | | |
| $\geq$ 21 days of paroxysmal cough | 14 | 32 | 58 (20–80) |
| $\geq$ 7 days of cough | 26 | 44 | 44 (7–67) |
| Younger siblings | | | |
| WHO definition | | | |
| $\geq$ 21 days of paroxysmal cough | 10 | 18 | 43 (−31 –76) |
| $\geq$ 7 days of cough | 11 | 10 | 37 (−40 –73) |
| Göteborg definition | | | |
| $\geq$ 21 days of paroxysmal cough | 10 | 26 | 61 (15–83) |
| $\geq$ 7 days of cough | 11 | 26 | 56 (9–81) |

## 12.5 Other approaches

Another approach to estimate $VE_S$ controlling for exposure to infection is to collect information on the number of contacts and then either use knowledge about the infection status of the actual contacts or information about prevalence of infection in the population from which the contacts are drawn. Yang et al (2009) developed a model to estimate the $VE_S$ of an HIV vaccine that used reported number of contacts and information on the prevalence of infection in the population. One of the study populations was an cohort of injecting drug users in Thailand. The contacts were drug injections with needles. Injections with shared needles were potentially infectious. The second study population was primarily men who have sex with men. The model allowed for errors in the reported number of contacts in each time interval. Case-contact studies are studies in which individuals exposed to a case are followed to find if they are infected or diseased. In both of these study types, there is no explicit transmission unit such as a household or a school.

## Problems

**12.1. Computing mumps $VE_{SAR}$**

(a) Table 12.4 shows the data from a family-based mumps vaccine efficacy study after an outbreak in Ashtabula County, Ohio, in 1982. In study 1, vaccine status was verified by the parents. In study 2, it was verified by the

provider. Compute the secondary attack rates and $VE_S$ with confidence intervals for both studies.
(b) Compare the estimates.

**Table 12.4.** Data from family-based mumps vaccine efficacy study in families of students with mumps illness in the sixth, seventh, and eighth grades in School A, Ashtabula County, Ohio, February 5 through April 23, 1982 (Kim-Farley et al 1982).

| | Study 1 | Study 2 |
|---|---|---|
| Case definition | Parotitis ≥2 days | Parotitis ≥2 days |
| Case finding | Parents | Parents |
| Vaccine status | Parents | Provider verified |
| Cases/exposed (vaccinated) | 4/36 | 2/30 |
| Cases/exposed (unvaccinated) | 32/74 | 30/69 |

### 12.2. Computing measles $VE_{SAR}$

(a) Table 12.5 contains data from a measles epidemic in Senegal 1994–1995 (Cisse et al 1999). Compute the estimates of the main $VE_S$, $VE_I$, and $VE_T$, the two stratified $VE_S$ and the two stratified $VE_I$. Compute their confidence intervals using the standard approach.
(b) Compare the main and the stratified estimates.

**Table 12.5.** Number of exposed susceptibles, secondary cases, and secondary attack rates (SAR) by vaccination status of the index case and the exposed susceptible children (Cisse et al 1999)

| | Exposed susceptibles and secondary cases | | | | | |
|---|---|---|---|---|---|---|
| | Vaccinated | | Unvaccinated | | Combined | |
| Index case | cases/exposed | SAR | cases/exposed | SAR | cases/exposed | SAR |
| Vaccinated | 6/83 | 0.07 | 3/17 | 0.18 | 9/100 | 0.09 |
| Unvaccinated | 41/374 | 0.11 | 47/124 | 0.38 | 88/498 | 0.18 |
| Total | 47 /457 | 0.10 | 50/141 | 0.35 | 97/598 | 0.16 |

### 12.3. Pertussis vaccine efficacy with different cutoffs

(a) Table 10.1 contains the number of secondary pertussis cases using four different follow-up cutoffs. Compute different $VE_{SARs}$ using different cutoffs.
(b) Discuss how and why the SARs and the $VE_{SAR}$ estimates change as the cutoff period increases.