Chapter 10
Household-Based Studies

10.1 Concepts of Household Studies

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 1930s in evaluating the efficacy of pertussis vaccines (Kendrick and Eldering 1939). In addition to evaluating vaccine efficacy, household studies have been used to learn about 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 (Storsaeter et al 1998). Longitudinal studies of pneumococcal carriage in households and schools have been used to estimate the acquisition and clearance rates for asymptomatic pneumococcal carriage.

The general idea of a transmission unit is that individuals make contact sufficient for transmission within it. Households are the most common form of transmission unit used in studies. 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, airplanes, day care centers, and workplaces, among others. Here we talk mostly about household studies, but many of the study designs and analyses are applicable with possibly slight modification to other transmission units as well. The term household is much easier for exposition than is "transmission unit".

Historically the use of household studies to evaluate vaccine effects 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. In recent years, the vaccine effect on the ability to transmit the infection in vaccinated infected people compared to unvaccinated infected people, VE$_I$, has gained attention. In contrast to protective effects, VE$_I$ generally needs contact and exposure to infection information for its evaluation. An additional measure of interest is the overall reduction in transmission if both the infective person and the susceptible person who make contact are vaccinated compared to if neither is vaccinated, VE$_T$. The analysis is often based on the 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 probability that an individual infects another person during some period of time. The secondary attack rate can be estimated from the 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 occurring 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).

Considering the estimates of VE based on the relative secondary attack rates, there are three main unstratified vaccine effects:

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

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

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

Equations (10.1) and (10.2) give the three main unstratified and three stratified vaccine effects conditional on exposure to infection data, conditional on exposure to infection. The vaccine efficacies in (10.1) and (10.2) could also be defined in terms of the relative transmission probabilities or transmission rates.

Despite being widespread for some infections such as pertussis, 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 primary analysis is generally based on one of the unconditional measures of vaccine efficacy, such as VE$_{S,IR}$ or VE$_{S,CI}$. 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. Household studies are also used in observational evaluation of

vaccines. 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 this and the following two chapters, we consider household 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 the introduction of pneumococcal vaccines. Similar studies are now being prepared by the MenAfriCar Consortium to anticipation of introducing the meningococcal A vaccine in the meningitis belt in Africa. 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. In many analyses of household studies, the households are assumed to be independent of one another, so that susceptible contacts are assumed 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. 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.

We introduce a few terms that are used. 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 often, but not necessarily, temporally the first, that is, the *primary case* in the household. 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*.

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 1920s and the first hopeful results were observed in the Faroe Islands in the early 1920s (Madsen 1933; Medical Research Council 1951). Most pertussis vaccines were based on killed whole cells until the 1980s. Concern about efficacy and adverse effects of whole cell pertussis vaccines resulted in some countries stopping to recommend their use. For example, Sweden

completely discontinued pertussis vaccination in 1979 because the efficacy seemed to be negligible (Trollfors and Rabo 1981). A new generation of acellular vaccines was developed as an alternative to the killed whole cell ones. In the 1980s and early 1990s, 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 whole cell and acellular 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 diphtheria and tetanus toxoids and given three to four times early in the first year of life. The vaccine combination 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 three- to four-month intervals, but after November 1935 at two-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 (2.3)). 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 of the exposure 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 (2.4).

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, information for each child 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 eight days or more. The pertussis vaccine studies were 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 estimators (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 biological 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, Prézios and Halloran (2003b) and Halloran, et al (2003b) analyzed the data to estimate not only VE_S but also VE_I and VE_T for pertussis. Prézios and Halloran (2003b) considered a number of different case definitions and the relation to estimated VE_S, VE_I, and VE_T. Halloran et al (2003b) considered different statistical methods for the secondary attack rate analysis (see Chapter 12.3) 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.

Préziosi and Halloran (2003b) 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", ie, the residential unit where individuals make privileged contacts and where random mixing is a reasonable assumption. The compound is the transmission unit of choice in some African rural settings (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 ≥7 days after that of the index case and less than a variable cutoff, specifically no cutoff, 56, 42, or 28 days.

Generally, when estimating protective efficacy, VE_S, from SARs, co-primaries are simply ignored in the analysis, entering as neither susceptibles nor infectives (Orenstein, et al 1988; Fine et al 1988). However, the particular interest here was in the effect of vaccine status on infectiousness of the index case. Because primaries and co-primaries often had different vaccine status, compounds with co-primaries were excluded from the analysis. Chu et al (2004) developed MCMC methods to estimate heterogeneous transmission with multiple infectives.

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 were 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 SARs using different cutoffs. The result of at least one biological confirmation criterion was available in over 97% of the suspected cases meeting the clinical def-

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

| | 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

inition. From the same study, Préziosi and Halloran (2003a) estimated the effect of pertussis vaccination on clinical severity, VE_P (Chapter 9).

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 months old 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 10 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 6710 children completed the trial, with 3358 in the vaccinated and 3352 in the unvaccinated group. In the vaccine group, there were 149 cases in 102,961 child-months at risk, and in the unvaccinated group, there were 687 cases

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

Vaccination Status	Home Exposure			Other Exposure			No Exposure History
	No. of Exposures	No. of Cases	Rate (%)	No. of Exposures	No. of Cases	Rate (%)	
Vaccinated	203	37	18.2	566	47	8.3	65
Unvaccinated	173	151	87.3	561	213	38.0	323

in 102,180 child-months at risk, a risk ratio of 1 to 4.6. The results give a $VE_{S,IR} = 1 - 1.45/6.72 = 0.78$ (95% CI, 0.74–0.82). Analysis of information on the exposures of children to pertussis was divided into two categories. First, 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 five 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 comprised 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 six 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 six 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 9515 cases were notified in 187,595 child-years at risk (5.1%) over the course of the study. Efficacy, $VE_{S,IR}$, based on the total number of cases for each year of birth was greater than 0.80. However,

Table 10.3 Secondary attack rates in home contacts according to age and vaccine group, England (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

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 7 days of the index case and secondary cases were those that occurred within about 42 days of the index case and at least 7 days after the index case. The efficacy was higher with a more severe case definition, reaching 71% 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 six 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 9242 households with 10,406 contacts, of whom 6436 (61.8%) developed pertussis at the same time or after symptom onset in the primary case. The 1520 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, although 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, although not significantly, in contacts exposed to vaccinated primary cases than in contacts exposed to unvaccinated primary cases. Thus, the two stratified VE_I estimates in equation (10.2) differed. 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 2003b). They

speculate that it could be due to household clustering of vaccine failures or false-positive diagnoses.

10.2.5 Sweden

After pertussis vaccination was discontinued 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 1990s.

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 Section 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 Section 15.3.2. 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

Prospective, 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, although they have been used for evaluating the effects of post-exposure prophylaxis of influenza antiviral agents. Household studies of influenza are particularly useful for studying

transmissibility and the serial interval. We present a number of household-based studies of influenza transmission for their historical significance and to promote future prospective household-based studies of influenza and other respiratory diseases. This sort of study of had essentially been discontinued. The novel influenza virus (H1N1) pandemic that started in 2009 has raised the consciousness about the important role of prospective household studies in estimating the transmissibility and the serial interval of influenza. We present the household-based studies of influenza antivirals to illustrate further methodological issues.

10.3.1 Seattle USA

Intensive surveillance of families with school-age children for influenza virus infections was conducted from 1975 to 1979 in Seattle, Washington, USA (Fox et al 1982b). The study followed the Virus Watch method that basically involves continuing virological surveillance of families. The Virus 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, biweekly 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, although there is no way to estimate how many. Between 9% (Group I) and 13% (Group II) of reported illnesses had no specimens collected, and between 26% (Group I) and 32% (Group II) of illnesses were recognized only because specimens were collected. Fox et al (1982a) 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, USA (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, USA, during the five-year period 1976–1981 (Monto et al 1985). Beginning in October 1976, recruitment over a three-month period resulted in 1000 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 eligibility. 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 preseason 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 were 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.

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, USA (Longini et al 1988)

Pre-Season Antibody Titer	Infection Status			Attack Rate
(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 et al 1964). The study had two primary objectives. 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 six-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 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 epidemiological data (Jordan et al 1958). Table

Table 10.6 Influenza attack rates by age during the Asian influenza pandemic of 1957 in Cleveland, Ohio, USA, as measured by virus isolation (Jordan et al 1958)

| Age Groups | | | Respiratory Illness | | | |
| | | | Test for Virus | | Virus Isolated | |
(Years)	No.	No.	No.	Percent	No.	Percent
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

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 ($\geq 38°C$) 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 (Halloran et al 2007a)

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

[*] 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 2007a). Table 10.7 contains a summary of some characteristics of the four studies. All four studies were household-based, multicenter, 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 years 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 the study drug (zanamivir) or placebo. Index cases did not receive antiviral therapy. Children under age five years 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 five days. Children under one 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, although 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 SAR_{jk}s, in principle, we can estimate the stratified antiviral efficacies, AVE_S, AVE_I, and 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 (Section 10.6.5). Second, asymptomatic infections in contacts were ascertained, so that pathogenicity and the effect of prophylaxis on pathogenicity, AVE_P, could be estimated. Third, each of the efficacies AVE_S, AVE_I, and AVE_T could be based on laboratory-confirmed influenza illness, $AVE_{.d}$, or simply laboratory-confirmed infection, $AVE_{.i}$, in the eligible contacts.

10.4 Measles Vaccination

Measles vaccines are generally much greater than 90% 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 when maternal antibodies are still present, vaccination will not be effective. On the other hand, if vaccinated too late, maternal

antibody protection will have waned, and the child could easily contract measles before being vaccinated. In the United States, vaccination against measles occurs between 12 and 15 months. However, in developing countries, this is often too late because exposure is more widespread. Considerable research has been directed at understanding the optimal age to vaccinate infants in developing countries. In the 1990s, 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 the 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 1566 children were vaccinated, with vaccine coverage of 81.6% of the resident target population. The analysis controlling for the level of exposure within the compound was nested in the randomized study.

Three measles outbreaks occurred during the study. In the first, 27 cases occurred between May and September 1988, then 161 cases between October 1988 and July 1989, and then 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 serological 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 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

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. HT = high titer. (Garenne et al 1993)

Group	Prospective study			Compound Exposure Study		
	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

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 Studies

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. Symptomatic 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 asymptomatic 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 and contained up to 23 of the serotypes. The first to be licensed in the United States was in 1977, with an improved version in 1983 (Plotkin and Plotkin 2008). Immunogenicity was not great, so a new generation of conjugate vaccines was developed based on purified polysaccharide joined to a harmless variety of diphtheria toxin. The conjugate pneumococcal vaccine was licensed in the United States 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 introducing the new vaccines, a series of household-based carriage studies was conducted 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) analyzed 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.

From September 2001 to May 2002, a further carriage study was conducted in Finland. It was the first longitudinal study of pneumococcal carriage to record serotype specific exposure to pneumococcal bacteria simultaneously with families and day care centers, the two most important mixing groups (Leino et al 2008). The

acquisition of pneumococcal carriage by day care attendees was strongly associated with previous exposure to a homologous serotypes in the day care center. In the 36 acquisitions with known exposure within the day care center or the family, the child had been exposed in the day care center in 35 cases and in the family in in 9 cases. The three day care centers were much larger than families, leading to the suggestion that the larger size and younger age of the children in the day care centers were about to main micro-epidemics better than the small families. The authors suggest that the day care centers serve as core populations to enhance pneumococcal transmission within the population as a whole. A child-to-child basic reproductive number was estimated as 1.4 (Hoti et al 2009) .

10.5.2 France

A five-month longitudinal study of three- to six-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). The mean time between consecutive swabs was 37 days (sd 15 days). During the observation period 9857 swabs were collected for serotyping. The 4488 three- to six-year old childen attending the schools represented 88% of the children in the area under study. Of these, 2445 (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 2006d). 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 (2006d) 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 3753 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
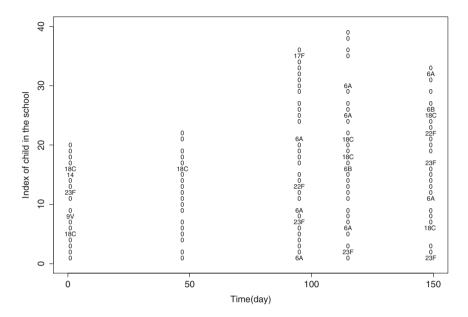
Fig. 10.1 Longitudinal data in a school participating in a pneumococcal carriage study. A "0" represents a sample in which no pneumococcal serotypes was detected. The other symbols represent the pneumococcal serotype abbreviation of the detected bacteria. Similar to Cauchemez et al (2006d).

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 four months old, then monthly up to one 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 1459 samples (92% of those planned) were collected from the 99 index children and 2865 from the other family members. Approximately 50% of the infants had acquired pneumococcal carriage by eight weeks of age. The point prevalence of pneumococcal carriage in the first five years was about 50% and declined after that 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 house-holds or other transmission units. A transmission unit is a place or social relation-ship within which individuals are assumed to make contact sufficient for transmis-sion. The concept of a contact sufficient for transmission is very broad and must be defined in each particular study. The transmission mode of an infectious agent deter-mines what types of contact are potentially infectious. Contacts can be defined be-tween two individuals, or an individual and a vector. Contacts can be defined within small transmission units, such as households. Within small transmission units, mix-ing is often assumed to be random. A small transmission unit can be defined as two individuals in a social relationship, such as a steady sexual partnership, or a house-hold 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 *mini-cohort* (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. Individuals living in the same household are likely to be more homogeneously exposed to infection. Comparing vaccinated and unvaccinated persons matched on household could be less prone to bias from differences in exposure to infection (Struchiner et al 1994). A small transmission unit can also be thought of as a *mini-community* 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 Ni-

akhar, Senegal, the SARs estimated in schools, in homes, and in huts differed (Cisse et al 1999). Kendrick and Eldering (1939) differentiated definite and indefinite exposures.

When collecting data on households, the identities and number of the people living in the household should be collected. Also, if there is interest in estimating transmission parameters or the secondary attack rate, it important to ascertain for each member of the household whether they were actually present in the household during the period of interest.

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, the case-ascertained design (Yang et al 2006), 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 the secondary attack rate would be estimated to be higher than if a random sample had been observed. 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, an example of the augmented study design (Section 10.7.4). 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 partici-

pant or a sibling, the nested study is subject to potential selection bias (Halloran and Struchiner 1995; Becker et al 2006).

A second issue is how cases within the household are 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, though not significantly, 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 (1988) 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 (Chapter 11).

If ascertainment of households is on an index case, then the duration of follow-up for each household needs to be determined, depending on the natural history of the infectious agent. 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 design. 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

In evaluating the effect of interventions, the assignment mechanism is key. We consider first that we are interested in estimating VE_S, VE_I, and 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, SAR_{10}, some of the households must have vaccinated primary cases and unvaccinated contacts. To estimate SAR_{11}, some of the households must have vaccinated primary cases and vaccinated contacts.

Table 10.9 Estimable antiviral efficacies from each of four household-based, household-randomized, influenza antiviral efficacy studies. AVE_I is not estimable from any of the studies alone (Halloran et al 2007a)

	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
$AVE_{S01/00} = 1 - \frac{SAR_{01}}{SAR_{00}}$	–	$AVE_{S01/00}$	–	$AVE_{S01/00}$
$AVE_{S11/10} = 1 - \frac{SAR_{11}}{SAR_{10}}$	–	–	$AVE_{S11/10}$	–
$AVE_{I11/01} = 1 - \frac{SAR_{11}}{SAR_{01}}$	–	–	–	–
$AVE_{I10/00} = 1 - \frac{SAR_{10}}{SAR_{00}}$	–	–	–	–
$AVE_T = 1 - \frac{SAR_{11}}{SAR_{00}}$	AVE_T	–	–	–

Most household-based studies of vaccine efficacy conducted up to now have been either observational studies or studies nested within individually randomized studies. In these studies, the allocation of vaccination within households generally is not under the control of the investigator. Theoretical and simulation studies have shown that to estimate VE_S, VE_I, and 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 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 $AVE_{S01/00} = 1 - SAR_{01}/SAR_{00}$ is estimable (AVE for antiviral efficacy). In the Osel I study, the index cases were all treated, and then all household contacts randomized to either drug or control. In Osel I, the other stratified $AVE_{S10/11} = 1 - SAR_{11}/SAR_{10}$ is estimable. In contrast, the Zan I study randomized everyone in a household, index cases and contacts, to either drug or control. In Zan I, $AVE_T = 1 - SAR_{11}/SAR_{00}$ is estimated. Without careful examination, one might believe that all three studies were estimating the same parameter, but there are not so subtle differences 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 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 AVE_I, although 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 (Section 12.3). If it is possible to control allocation of vaccination or other interven-

tion 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 VE_I than to estimate VE_S. 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 VE_I. However, if VE_S is high, 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, HIV, or the novel influenza (H1N1) pandemic, 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. Case-contact studies are studies in which individuals exposed to a case are followed to find if they are infected or diseased. In this type of study, there is no explicit transmission unit such as a household or a school.

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 VE_S conditioning on potential exposure to infection is to assemble a cohort of susceptibles. The study then follows the susceptibles and collects information on their contacts with infectives or potential infectives. One can use either information about the infection status of the actual contacts or information about prevalence of infection and contact structure in the population from which the contacts are drawn. 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 data required are infection outcome, number of potentially infective contacts, and covariate status, for example, vaccination status, for each person in the study. Yang et al (2009a) 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.

10.7.4 Augmented vaccine studies

It is possible to design studies prospectively that intentionally make use of multi-level 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 VE_S, although estimation of VE_I is also possible. The individual recruitment and randomization are 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 this type of study 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 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 transmission probabilities. Similar to the community-randomized trial 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.

Problems

10.1. (a) Describe the main differences in the design of prospective versus case-ascertained household studies.
(b) What are the advantages and disadvantages of the two approaches?
(c) What are the differences in the potential sources of bias?

10.2. (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. (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?

10.4. Consider designing a household-based study of a new vaccine targeted to children under 6 years old. The study will include only households with at least two children under 6 years old. Using a placebo as control is ethical for this vaccine. What vaccine efficacy measures will be estimable if you randomize by household? by individual?