13

Assessing Indirect, Total and Overall Effects

13.1 Study Designs for Dependent Happenings

Due to the dependent happenings in infectious diseases (Ross 1916), widespread vaccination in a population can reduce transmission and produce indirect protective effects, even in unvaccinated individuals The public health importance of a vaccine is related to the direct protection of the vaccinated individuals as well as the indirect protection conferred by increased herd immunity at the population level. In recent years, interest in estimating the indirect and overall effects of vaccination programs has increased. Most often, the effects have been evaluated using surveillance data by comparing the incidence before and after implementation of a vaccination strategy in a population. In some cases, dramatic effects have been observed such as with pneumococcal vaccines (Musher 2006). Up until now, planned, prospective community-randomized studies to evaluate indirect, total, and overall effects of vaccination strategies are rare. However, interest in implementing such studies, either pre- or post-licensure is increasing. Though mathematical models offer useful guidance on examining potential population effects of vaccination strategies (Chapters 4 and 5), they cannot replace data from an actual study when such a study is feasible.

Struchiner et al (1990) and Halloran and Struchiner (1991) developed a conceptual framework for four classes of study designs to evaluate the direct, indirect, total, and overall effects of interventions. In Chapter 2 we introduced the general concepts of direct, indirect, total and overall effects of vaccination and the four basic study designs to evaluate them. In this chapter, we present the definitions in the general potential outcomes approach to causal inference, initially using a heuristic approach. Throughout this chapter we distinguish two levels of intervention, vaccination of individuals and vaccination strategies, allocations or programs within populations. We discuss some of the less formal approaches to assessing indirect and overall effects, their advantages and disadvantages. We then present approaches to community-randomized studies that could be used for more formal estimation and inference of indirect, total, and overall effects. We discuss some further considerations when

designing interventions to evaluate the effects. We integrate studies of herd immunity with the literature on group-randomized studies. Then we consider basic designs, approaches to randomization, sample size determination, and general considerations of analysis. Finally we formally define causal estimands of direct, indirect, total, and overall effects and their estimators for group randomized studies.

13.1.1 Definitions and Study Designs

Following Halloran and Struchiner (1991), the *direct effect* of vaccination in an individual is the difference between the outcome in the individual receiving the vaccine and what the outcome would have been if the individual had not been vaccinated, all other things being equal. This definition of a direct effect corresponds to the notion of potential outcomes in causal inference in that it is defined for the unobservable difference between the response in the observed person and what it would have been in the same person without the intervention. An example of a direct effect is the reduction in the probability of becoming infected that results from being vaccinated, given exposure to infection.

The *indirect effect* of a vaccination program or strategy on an individual is the difference between what the outcome is in the individual *not* being vaccinated in a community with the vaccination program and what the outcome would have been in the individual, again not begin vaccinated, but in a comparable community with no vaccination program. It is, then, the effect of the vaccination program on an individual who personally was not vaccinated. The combined *total effect* in an individual of being vaccinated and there being a vaccination program is the difference between the outcome in the individual being vaccinated in a community with the vaccination program and what the outcome would be if the individual were not vaccinated and the community did not have the vaccination program. The total effect, then, is the effect of the vaccination program combined with the effect of the person having been vaccinated. The *overall effect* of a vaccination program is the difference in the outcome in an average individual in a community with the vaccination program compared to an average individual in a comparable population with no vaccination program.

A simple indirect effect is the reduction in the probability per unit time of becoming infected that results from reduced exposure to infection consequent to a mass immunization program. Thus, an unvaccinated person in a population experiences a changed hazard or incidence compared with what it would have been if the commmunity had had no immunization program. The analogous total effect would be the effect experienced by a vaccinated person who has both the benefits of being vaccinated and the indirect effect of the reduced transmission. The overall effect would be the weighted average of the reduction in incidence in the vaccinated and unvaccinated individuals compared to if there were no immunization program.

These effects can be defined more generally, by allowing that vaccination occur in the the comparison population as well, however with a different level of coverage or different allocation strategy. The definitions also apply to subpopulations of interest with comparison populations, for example, schoolage children only. The indirect, total and overall effects are defined within the context of a particular intervention program or allocation strategy. For example, one would expect that the indirect effects of vaccinating 30% of the population might differ from the indirect effects of vaccinating 60% of the population compared to no vaccination. Common to these effects is the need to imagine a community in which vaccination had not taken place or with an alternative vaccination strategy.

The four different kinds of effects of vaccination motivated the definition of broad categories of study designs (Struchiner et al 1990). based on different pairs of comparison populations and subpopulations, according to whether the studies measure direct, indirect, total, or overall effects as shown in Figure 2.3. In the simple case of having only two comparison populations, the study designs for dependent happenings are analogous to studies that compare pre- and post-implementation of a vaccination strategy in a population. However, since the community level effect is of interest, for statistical inference, one generally will need several communities in which the intervention takes place and several comparison communities. If the allocation of the vaccination program to the communities is randomized, then the study becomes a group- or cluster randomized design as discussed in Section 13.3.

The indirect effects of the vaccine given a particular allocation of vaccination is then the comparison of the incidence or other outcome of interest in the unvaccinated people in the A communities compared to the unvaccinated people in the control B communities. These comparisons are called designs type IIA. The indirect effectiveness measures are designated $\mathrm{VE}_{I\,I\,A}$. The total effects of the combination of being vaccinated and the allocation is the outcome in the vaccinated people in the A communities compared to that of the unvaccinated people in the unvaccinated B communities. These comparisons are called designs type IIB, and the total effectiveness measures are designated $\mathrm{VE}_{I\,I\,B}$. The overall effectiveness of the vaccine and allocation compare the average outcomes in the vaccinated communities with those of the control communities. These comparisons are called designs type III, and the overall effectiveness measures are designated $\mathrm{VE}_{I\,I\,I}$.

These study designs are quite general. They do not specify the outcome measure, the parameter of effect, temporal aspects, sampling methods, or methods of analysis. In addition, each of these designs makes a comparison between comparable populations, in that, in the absence of an effect, the outcomes of the compared populations could be expected to be similar. In particular, in infectious diseases, one needs to emphasize the need for comparability of exposure to infection. That is, if there were no intervention program, the individuals in the comparison communities would be exposed comparably to infection. This is often not the case, especially when several communities are

included in the study, in which case the design can incorporate matching or stratification as described in Section 13.5.

Table 2.2 contains examples of the $VE_{I/A}$, $VE_{I/B}$, and $VE_{I/I}$ based on the usual unconditional measures incidence rate, hazard rate, and cumulative incidence. Many other measures could be used, including average age of infection or the basic reproductive number, R_0.

13.2 Observational Studies

13.2.1 Pre- and post-vaccination comparisons

A common aproach to estimating indirect, total, and overall effects of introducing a new vaccination program to a population is to compare the pre-vaccination with the post-vaccination incidence. These comparisons depend on good data on cases of the illness of interest and some method to determine the denominators. To determine indirect or total effects, one also needs to know the vaccination status of the reported cases. One might also want to know the level of vaccine coverage and the age-appropriate and age-specific vaccine uptake.

Comparisons or pre- and post-vaccination outcomes include comparison of incidence or attack rates before and after introduction of vaccination (overall effects), possibly also stratified by vaccine status (indirect and total effects), reduction in incidence greater than vaccine coverage, reduction in incidence in age groups that did not receive vaccination, change in the age distribution of disease, and increased prevalence of colonization and disease by nonvaccine strains (Table 13.1).

Table 13.1. Comparisons pre- and post-introduction of a vaccination strategy to estimate indirect, total, or overall effects

Comparison
Change in incidence or attack rates in target population (overall effects), possibly stratified by vaccination status (indirect and total effects)
Reduction in incidence in age groups that did not receive vaccination strategy (indirect)
Reduction in incidence greater than vaccine coverage (overall)
Change in age distribution of disease
Increased prevalence of colonization and disease by nonvaccine strains

If the reduction in overall incidence, including both the vaccinated and unvaccinated individuals combined, is higher than the level of coverage, there is a strong indication of indirect effects of vaccination. Thus, even with a

100 percent efficacious vaccine, if coverage were 60%, one would not expect a greater than 60% reduction in incidence if there were no indirect effects. Thus, an observed 80% reduction in incidence would be evidence for indirect benefits of vaccination. Another indication that herd immunity is playing a role is a reduction in incidence in age groups that are too young or too old to be in the age group targeted by the strategy. Reduction of incidence in these groups is evidence of a purely indirect effect of the vaccination program. The mean and median age of first infection will generally increase as transmission is reduced, since it will take on average longer for a person (child) to be exposed. As transmission is reduced, incidence in all age groups may decrease, but the relative proportion of cases in the older age groups could increase. The change in age of first infection is also an indication that the reproductive number R is changing. In cases of infectious agents with many circulating strains, only some of which are contained in the vaccine, such as *Streptococcus pneumoniae*, there is interest in whether the prevalence of colonization and incidence of disease due to the nonvaccine strains will increase as the prevalence of vaccine strains is reduced.

If an observed change in outcome, such as a change in incidence rate, is to be attributed to the vaccination strategy when comparing only the pre- and post-vaccination situation, one needs to make an assumption of minimal secular trends. When the pre- and post-vaccination differences are small, and one is comparing only one pre- to one post-vaccination population, one cannot be sure that some other cause than vaccination is not responsible for any observed changes. For example, change in sanitation or simple cyclical variation of the infection rates might decrease incidence. Also, if the duration of observation is short, for instance, a comparison of influenza one year and then the following year in which a vaccination campaign was done, one cannot be sure that the second year was not simply a milder year.

Another approach to estimating effects of widespread vaccination is to compare data from different regions with different levels of coverage. Ali et al (2005) re-analyzed an individually randomized trial of cholera vaccine by comparing incidence in areas with different the coverage levels (Section 13.2.6). However, the level of vaccine uptake may be related to other factors related to the level of incidence. Thus the estimates of indirect effects could be confounded, unless the coverage levels were randomized.

13.2.2 Pertussis in Niakhar, Senegal

Préziosi et al (2002) studied pertussis in a prospective cohort of children in rural Niakhar, Senegal over a 13-year period comprising time before and after introduction of a pertussis vaccination program. Children under age 15 years who were residents of the Niakhar study area were followed prospectively between January 1984 and December 1996 for the occurrence of pertussis. (See Chapter 10.2.3 for further details.) From 1980 to 1985, sporadic immunizations were performed, reaching fewer than five percent of the children.

From November 1986 to January 1987, Senegalese authorities conducted Expanded Program of Immunization (EPI) mass immunization campaigns targeting children under age 5 years. After August 1987, infants were immunized by monthly visits of the EPI mobile teams, with rigorous record keeping. From 1987 to 1989, children received whole cell pertussis vaccine as part of DTP-IPV at approximately 3, 5, and 10 months of age. From 1990 to 1996, clinical trials of the relative efficacy of whole cell pertussis and acellular pertussis vaccines were conducted with vaccination at 2, 4, and 6 months of age. A child who had received three doses of pertussis vaccine regardless of vaccine type was considered to be fully immunized.

Vaccine uptake was measured by the number of children who received three doses of pertussis vaccine before the end of the calendar year of their first birthday, divided by the number of live births. Vaccine coverage was evaluated by the number of fully immunized children resident on December 31st of the year, divided by the corresponding number of residents per age group. Pertussis incidence rates were calculated using a person-time incidence density approach. One unit was added to each monthly total to avoid null values and a moving average over five months was used to smooth variations.

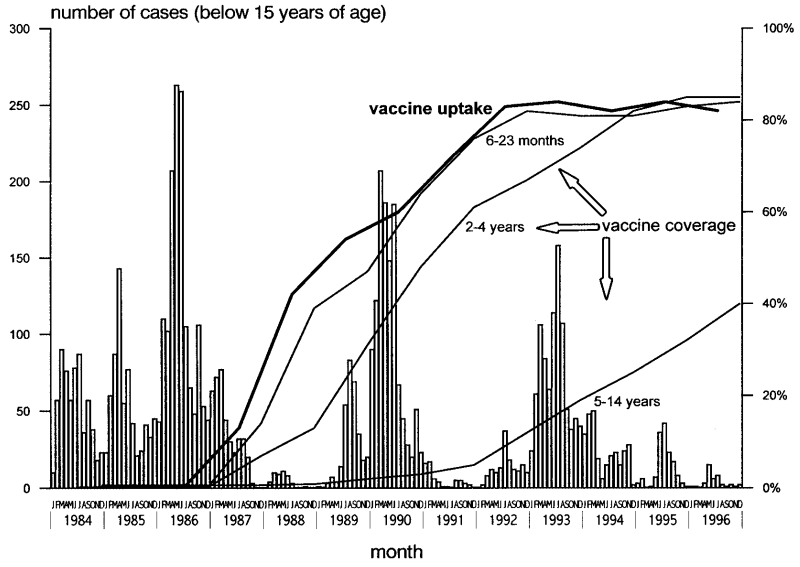

Fig. 13.1. Pertussis cases per month, vaccine uptake, and age-specific vaccine coverage per year, Niakhar, Senegal, 1984–1991 (Préziosi et al 2002)

EPI vaccine uptake rose from 13 percent in 1986 to 72 percent in 1990, and finally reached a level of 82 to 84 percent (Figure 13.1). High vaccine coverage

(>80 percent) was achieved in the youngest age group (6 months to 1 year) by 1991, but remained relatively low at 40 percent in the 5 to 14 age group even up to 1996. Pertussis was endemic, with annual peaks and epidemics every 3 to 4 years, centered on 1986, 1990, and 1993. Both the number of cases between epidemics and the magnitude of the epidemic peaks decreased. (Figure 13.1).

From late 1987 onward, the number of cases reported dropped between epidemic years (Figure 13.2). The decrease in incidence was observed in every age group, but especially in children under age 5 years. The greatest decline was in children under age 2 years. The declining trend was with a time lag according to age group . The overall effect of the pertussis vaccination program as measured by the reduction in incidence in the 0 to 14 year olds between the first and third epidemic peak was $\mathrm{VE}_{/\,/\,/} = (127.3 - 68.9)/127.3 = 0.46$ (Table 13.2). The most dramatic decline was for the children aged 6 to 23 months, where the reduction in incidence, or overall effectiveness of the program was $\mathrm{VE}_{/\,/\,/} = (170.5 - 36.3)/170.5 = 0.79$. The indirect and total effects are not estimable from the data in Table 13.2 since the incidence rates by vaccine status are not given. The median age of pertussis cases rose steadily from 4.1 years in 1986 to 5.3 years in 1990 and 6.2 years in 1993 (Figure 13.2).

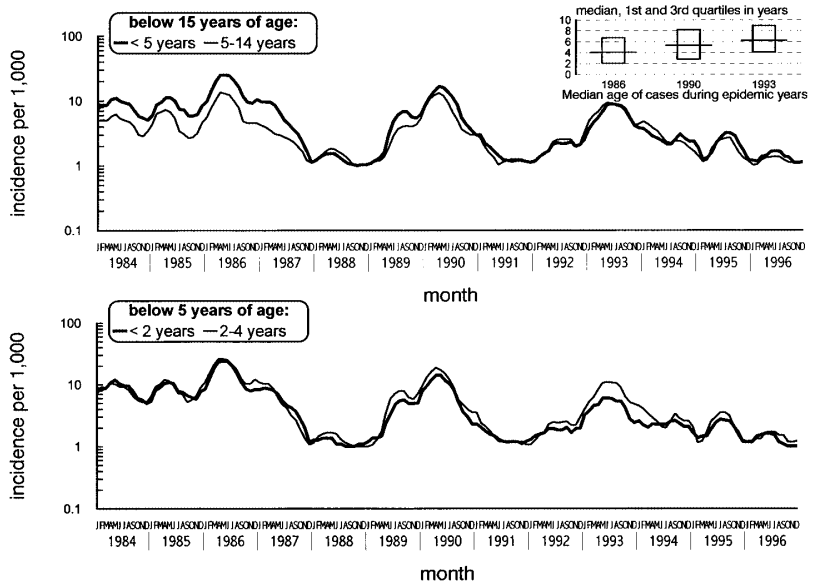

Fig. 13.2. Age-specific incidence rates of pertussis per period, Niakhar, Senegal, 1984–1996. (Préziosi et al 2002)

Table 13.2. Pertussis case distribution and incidence per age, during epidemic years, Niakhar, Senegal, 1984–1996 (Préziosi et al 2002)

Age	First outbreak (1986)				Second outbreak (1990)				Third outbreak (1993)			
	Cases		No. of PYR	Incidence/ 1,000 PYR	Cases		No. of PYR	Incidence/ 1,000 PYR	Cases		No. of PYR	Incidence/ 1,000 PYR
	No.	%			No.	%			No.	%		
0-5 mo	97	7	582	166.6	68	6	557	122.1	38	4	575	66.1
6-23 mo	246	18	1,443	170.5	144	12	1,700	84.7	58	7	1,598	36.3
2-4 yr	492	35	2,530	194.5	348	30	2,850	122.1	241	27	2,969	81.2
5-14 yr	570	40	6,481	88.0	612	52	7,422	82.5	555	62	7,811	71.1
Total	1,405	100	11,036	127.3	1,172	100	12,529	93.5	892	100	12,953	68.9

13.2.3 Pertussis in England and Wales

Miller and Gay (1997) discuss the effect of vaccination on pertussis epidemiology in England and Wales. Vaccine uptake dropped dramatically after 1974 (Figure 13.3) followed by a resurgence of pertussis cases in 1978. There has been considerable discussion in the literature about whether pertussis vaccination actually alters the transmission of pertussis in a population (See Chapter 12). Although there had been speculation that the drop in cases before 1974 had been due to improved social conditions, the steep increase with decreasing uptake is evidence that the drop in cases before 1974 was due to vaccination. Miller and Gay suggest that the decline in incidence is greater than would be expected given the low protective efficacy estimates of pertussis vaccination (see Chapter 10). They present modeling results assuming a $\mathrm{VE}_I = 0.80$ and graphically compare the number of cases from the transmission model with the observed number of cases in childern less than 3 months of age, too young to have received the vaccine. The results are consistent with pertussis vaccination lowering transmission and therefore likely there are indirect effects of vaccination. Also the age distribution of pertussis cases increased in England and Wales, though not as much as predicted by the transmission model. Underdiagnosis of pertussis cases in adolescents could result in underreporting of cases in that age group.

Taranger et al (2001) considered that mass vaccination of children with pertussis toxoid decreased incidence in both vaccinated and nonvaccinated persons in Sweden.

13.2.4 Pneumococcal vaccine in Alaska

Hennessy et al (2005) evaluated invasive pneumococcal diseases (IPD), antimicrobial resistance, and nasopharyngeal colonization before and after introduction of pneumococcal conjugate vaccine (PCV7) in Alaksa Natives. On January 1, 2001 PCV7 was introduced into the childhood vaccination schedule for all Alaskan children. Population-based surveillance for IPD among persons

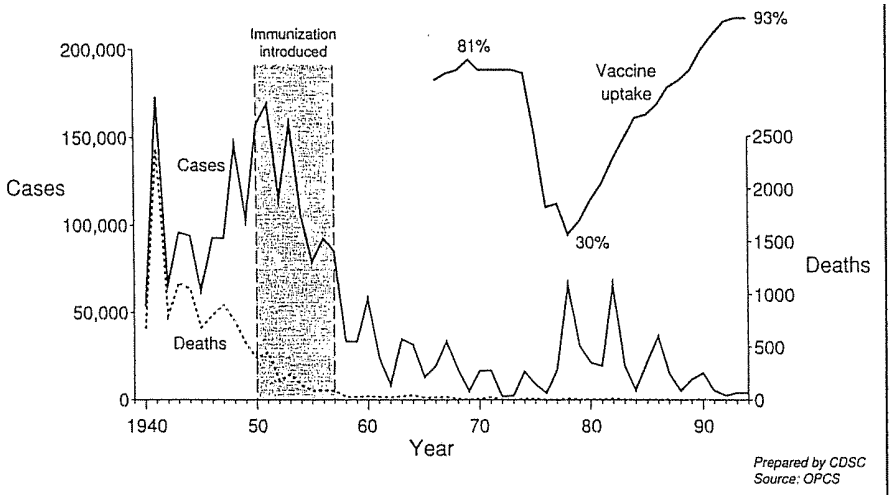

Fig. 13.3. Pertussis (whooping cough) notifications: cases and deaths. England and Wales 1940-94. (Miller and Gay 1997)

of all races throughout Alaska was conducted by the CDC Arctic Investigations Program, in place since 1986. Hennessy et al (2005) used the statewide surveillance for IPD to compare rates of disease in the six years prior to routine use of PCV7 (1995-2000) with disease rates in the three years after PCV7 use (2001-2003).

From October 1, 2001 to September 30, 2003 the proportion of 3-15 months old Alaska Native children who were age appropriately vaccinated with PCV7 increased from 51.9% to 73.2%. The proportion of 16-27 months old children with 4 or more PCV7 doses increased from 0 to 57.7%. By September 30, 2003, 95% of 19-35 months old Alaska Native children had received at least one dose of PCV7. From 1995 to 2003 a total of 1,113 case of IPD were reported in Alaska. Isolates were available on 90% of the cases. Table 13.3 shows the before and after rates and number of cases of IPD. The overall effectiveness against all serotypes in Alaska Native children <2 years was $VE_{///} = (403 - 142)/403 = 0.65$ and in non-Natives was $VE_{///} = (133 - 51)/133 = 0.62$, both of which were found to be statistically significant, ignoring that comparison is just before and after in one population. In children aged 2–4 years the overall effectiveness against all serotypes in Alaska Native children was $(73.9 - 12.7)/73.9 = 0.83$ but was just 0.10 in non-Natives. Most of the dramatic decline was in the vaccine serotypes. Overall effectiveness against PCV7 serotypes among children >2 years for Alaska Natives was $(275 - 25)/275 = 0.91$ and in non-Natives $(101 - 20)/101 = 0.80$.

Colonization studies were also conducted from 1998 to 2003 community-wide in eight rural Alaska villages and in urban clinics from 2000 to 2003 in children aged 3-59 months. The proportion of persons colonized with *S. pneu-*

Table 13.3. Rates (per 100,000) of invasive *Streptococcus pneumoniae* by time period, age group, race and vaccine serotype, Alaska, 1995–2003 (Hennessy et al 2005)

Age	Alaska Natives			Non-Alaska Native		
group (years)	1995–2000 rate(number)	2001–2003 rate(number)	*P*-value	1995–2000 rate(number)	2001–2003 rate(number)	*P*-value
Conjugate vaccine serotypes (4, 6B, 9V, 14, 18C, 19F, 23F)						
<2	275.3 (84)	24.7 (4)	<.001	101.3 (86)	20.0 (9)	<.001
2–4	47.0 (21)	0 (0)	<.001	13.6 (17)	7.5 (5)	.247
5–17	5.9 (12)	0.9 (1)	.035	1.0 (6)	2.5 (8)	.095
18–44	6.1 (6)	5.7 (8)	.909	4.3 (52)	1.09 (7)	.792
≥45	15.1 (23)	13.6 (11)	.792	11.4 (102)	7.4 (35)	.023
Non-conjugate vaccine serotypes						
<2	95.1 (29)	105.0 (17)	.738	23.6 (20)	28.8 (13)	.568
2–4	13.4 (6)	8.4 (2)	.610	4.0 (5)	7.5 (5)	.333
5–17	7.8 (16)	5.5 (6)	.484	2.6 (16)	1.5 (5)	.307
18–44	16.6 (44)	17.8 (25)	.779	3.6 (43)	2.8 (18)	.403
≥45	32.9 (50)	54.6 (44)	.016	10.4 (93)	7.0 (33)	.043
All cases (including unknown serotypes)						
<2	403.2 (123)	142.0 (23)	<.001	133.1 (113)	51.0 (23)	<.001
2–4	73.9 (33)	12.7 (3)	<.001	18.4 (23)	16.6 (11)	.792
5–17	15.2 (31)	8.3 (9)	.103	3.9 (24)	4.6 (15)	.616
18–44	25.3 (67)	24.9 (35)	.947	9.2 (111)	4.7 (30)	<.001
≥45	57.9 (88)	75.7 (61)	.112	23.5 (210)	16.9 (80)	.010

moniae of PCV7 serotypes declined substantially after PCV7 introduction. Decreased vaccine-type colonization and invasive disease in adults demonstrate indirect effects. Although not all denominators are given, Hennessy et al (2005) estimate that in ≥5 year olds, who were not eligible to receive PCV7, 41 cases of vaccine type IPD (95% CI 20-64 cases) were indirectly prevented by PCV7 introduction.

13.2.5 Meningococcal vaccine in the United Kingdom

In November 1999, the United Kingdom introduced routine meningococcal serogroup C vaccination for infants. The vaccine was also offered to everyone aged under 18 years in a phased catch-up program. Adolescents were vaccinated first and the program was completed by the end of 2000. Ramsay et al (2003) compared cases in unvaccinated children from each age group in the period from July 1, 2001 to June 30, 2002 with those in the same age groups for the period from July 1, 1998 to June 30, 1999. The denominator was mid-1999 population estimates from the Office of National Statistics for the age group, adjusted for the proportion of each cohort vaccinated. The

cases were identified at the Public Health Laboratory Service by confirmation of serogroup C disease. They investigated the vaccination history of all such identified cases. They computed vaccination coverage from data from immunization coordinators and departments of child health in England. They identified a total of 37 cases in the 2001-2002 period in the cohorts targeted for vaccination, eight in vaccinated children and 29 in unvaccinated children.

Table 13.4. Attack rate of confirmed meningococcal serogroup C infection in unvaccinated children before and after the launch of the vaccination campaign (Ramsay et al 2003)

	July 1998–June 1999				July 1998–June 1999				
Cohort	Date of birth	Cases	Population	AR per 100,000 (95% CI)	Date of birth	Cases	Est coverage (%)	Est. pop	AR per 100,000 (95% CI)	Indirect effect (95% CI)
Adolescent		96	1,818,034	5.28 (4.2, 6.3)		11	66	614,110	1.79 (0.7, 2.8)	66 (37, 82)
Grades 7–10		141	2,546,938	5.54 (4.6, 6.4)		4	86	359,118	1.11 (0.02, 2.2)	80 (46, 93)
Grades 1–6		76	3,911,606	1.94 (1.5, 2.4)		5	87	498,068	1.00 (0.1, 0.9)	48 (−28, 79)
Preschool		81	2,055,120	3.94 (3.1, 4.8)		6	76	501,449	1.20 (0.2, 2.2)	70 (30, 87)
Toddlers		41	601,045	6.82 (4.7, 8.9)		2	84	97,369	2.05 (−0.7, 4.9)	70 (−24, 93)
Infants		24	320,562	7.49 (1.5, 10.5)		1	80	64,112	1.56 (−1.5, 4.6)	79 (−54, 97)
Total		459	11,235,305	4.08 (3.7, 4.5)		29		2,134,226	1.36 (0.86, 1.85)	67 (52, 77)

Table 13.4 contains the number of cases in the unvaccinated children before and after launch of the vaccination campaign. The estimated indirect effect in children based on the attack rate over all age groups is $\mathrm{VE}_{I/A} = (4.08 - 1.36)/4.08 = 0.67$ with a 95% CI (0.52, 0.77) with a range of 0.48 to 0.80 in the different age groups. Using a denominator of 9,119,078 for the eight vaccinated cases for an attack rate of 0.09/100,0000, the estimated direct protective efficacy of the vaccine is $(1.36 - 0.09)/1.36 = 0.93$, with a 95% CI (0.86, 0.97).

13.2.6 Cholera vaccine in Bangladesh

Ali et al (2005) re-analyzed data from a large-scale, double masked, individually randomized field trial of killed whole-cell cholera vaccines given orally, either with or without cholera toxin B subunit in Bangladesh to ascertain whether there was evidence of indirect as well as direct vaccine protection of individuals. The trial was done in the Matlab field area of the International Centre for Diarrhoeal Disease Research, Bangladesh (ICDDR,B): Centre for Health and Population Research during the 1980s (Clemens et al 1990). All children aged 2 to 15 years and women older than 15 years were randomized to receive either one of the cholera vaccines or *Escherichia coli* K12 placebo. The main objective of the original trial was to assess whether receipt of three doses of vaccine was associated with lower incidence of cholera than that observed after receipt of three doses of placebo. At one year of follow-up, protective efficacy was 62% for B subunit-killed whole-cell oral cholera vaccine and 53% for killed whole-cell only oral cholera vaccine. The re-analysis to assess indirect

effects was motivated by the lack of enthusiasm for introducing the vaccine in populations with endemic cholera because of the moderate direct protective effects.

A bari in Bangladesh is a patrilinearly-related household living in clusters. Ali et al (2005) chose the bari as the unit of analysis because they are geographically discrete and because there may be transmission within these units. A total of 6,423 baris were included in the analysis, with the median number of individuals in a bari eligible for the trial being 17 (interquartile range 7-26). The analysis was restricted to the first year of follow-up to have a more stable population and minimize the effects of migration. Level of vaccine coverage was defined as the number of vaccinated individuals divided by the number of people who were eligible for participation in the trial by age and sex criteria. Then because the coverage of nearby baris might affect the risk of cholera of bari residents, the vaccine coverage of the bari was defined as the coverage of bari residents and those living within a 500 meter radius according to a geographic information system mapping.

For the indirect and total effects, models based on generalized estimating equations with a logit link and exchangeable correlation matrix including potential confounding variables were used (see Section 13.7). The occurrence of cholera in each analyzed individual was the dependent variable. The vaccine status of the individual (vaccine or placebo) and level of vaccine coverage of the individual's bari coded as a percent, as well as other potential confounding variables were fitted as independent variables. This analysis did not backtransform to the relative risk scale as in Préziosi and Halloran (2003) (Chapter 12), but assessed vaccine effects based on the odds ratios by exponentiating the coefficients of the logit model.

Table 13.5. Risk of cholera in placebo and recipients of killed oral cholera vaccines, by level of coverage of the bari during the first year of follow-up (Ali et al 2005)

Level of vaccine coverage	Target population	Vaccine recipients			Placebo recipients			Protective Efficacy (95% CI)
		N	Cases	Risk per 1000 population	N	Cases	Risk per 1000 population	
< 28%	24,954	5627	15	2.67	2852	20	7.01	62 (23 to 82)
28-35%	25,059	8883	22	2.48	4429	26	5.87	58 (23 to 77)
36-40%	24,583	10772	17	1.58	5503	26	4.72	67 (36 to 83)
41-50%	24,159	11513	26	2.26	5801	27	4.65	52(14 to 73)
> 51%	22,394	12541	16	1.28	6082	9	1.48	14(-111 to 64)
Total	121,149	49,336	96	1.94	24,667	108	4.37	56(41 to 67)

Table 13.5 presents a summary of the data divided into quintiles by level of coverage of the baris and the protective efficacy for each quintile. The risk of cholera in recipients of two or more doses of either vaccine or placebo is inversely related to the level of vaccine coverage of the bari. The trend is statistically significant in placebo recipients (Spearman's correlation coefficient -1.00, $p = 0.02$), but not in vaccine recipients (-0.90, $p = 0.08$). Three

analyses were done using the generalized estimating equations, one using all recipients with ≥2 doses (overall effect), one for those with vaccine (total effect), and one for those with placebo (indirect effect). The odds ratios for the level of cholera vaccine coverage of the bari were 0.97 (95% CI 0.96-0.98), 0.98 (0.96-1.00), 0.96 (0.94-0.98). respectively. Thus, there was a significant gradient by level of coverage for the overall effect and the indirect effect, with a borderline significant gradient for the total effect. They conclude there was an inverse, monotonic trend for the relation between the level of vaccine coverage in a residential cluster and the incidence of cholera in individual vaccine recipients or placebo recipients residing in the cluster after controlling for potential confounding variables.

13.2.7 Drawbacks of unplanned evaluation

Often evaluation of indirect, total or overall effects based on the pre- and post-vaccination surveillance data can provide good evidence, at least of the overall effects. However, if the change is to be attributed to the vaccination program, one must assume there are no major secular trends. Without any planned studies, the indirect effects of Hib vaccination took quite a while to estimate (Moulton et al 2000). A particular example of the difficulty of using unplanned studies, or studies based on comparing just one or two populations is influenza (Halloran and Longini 2006). Attempts have been made before to demonstrate the community-wide effectiveness of vaccinating school children against influenza. Just before the epidemic in 1968, Arnold Monto and colleagues vaccinated 85 percent of the school-age children in Tecumseh, Michigan, against influenza, resulting in a 67 percent decrease in the influenza-like illness attack rate in Tecumseh compared with neighboring Adrian (Monto et al 1969). In an ongoing community vaccination study in Central Texas with LAIV, Paul Glezen and colleagues are attempting to demonstrate that vaccinating school children reduces incidence of influenza-like illness in adults (Piedra et al 2007). Although these studies are rigorous, they each have only one or two comparison community.

A larger scale study with numerous comparison communities is needed to gather convincing data to counter any remaining scepticism. A study in several schools in the former Soviet Union used a nonspecific outcome as well, so the results are difficult to interpret (Monto et al 1993). A compelling example of the need to plan evaluation prospectively is provide by the Japanese national vaccination strategy, which for over two decades until 1987, was targetted at school children precisely to reduce epidemic influenza. A retrospective reassessment suggesting that the Japanese strategy reduced excess deaths among elderly adults (Reichert et al 2001) is open to criticism because it is based on non-specific mortality data over time. The time trends could result from factors not related to influenza vaccination. More recently, the province of Ontario, Canada, has been promoting wide spread vaccination for all age groups. The analysis of the Ontario experience suffers from

similar weaknesses as the Japanese. A review of 14 studies concluded that further evidence is needed of the indirect effects of influenza vaccination in children (Jordan 2005). King et al (2006) tried to demonstrate that school-based influenza vaccination reduced spread of influenza in households and communites, but used an influenza-like illness outcome, not influenza. The use of a nonspecific case definition compounds the difficulty of evaluating the indirect effects of influenza vaccination strategies.

13.3 Group-randomized Studies

Ideally, to evaluate indirect, total, and overall effects of a vaccination strategy, one would randomize several communities to receive the vaccination strategy of interest and several communities to serve as controls, and then the outcomes in the intervention communities would be compared with those of the controls. Most commonly, the luxury of conducting a prospectively designed study of a vaccination strategy in multiple groups or populations to estimate indirect, total, and overall effects will not be an option. The more feasible approach will often be to plan well for a comparison of the pre- and post-implementation incidence in the relevant populations as the studies described in the previous section. Despite the increasing interest in using group-randomized studies to evaluate population-level effects of vaccination, few actual studies have been conducted up to now. However, prospectively designed community-randomized studies may become more common in the future.

Community- or group randomized studies are those in which the intervention, or intervention strategy is randomized to groups of individuals. With vaccines, one can randomize the vaccination strategy at the group level and further randomize vaccination or control at the individual level, if desired. There is an extensive literature on group-randomized designs (Murray 1998). Often cluster-randomized studies are conducted because it is not feasible to allocate the intervention individually, even though the effects on the individuals are of interest. Occasionally, vaccination studies use a group-randomized design even when the direct protective effects are of interest because of practical or ethical consideration. Group-randomized designs are occasionally used in households where the parents or other household members might be unwilling to do a discordant, individual randomization. We discuss group-randomized studies here primarily for our interest in measuring indirect, total, or overall effects.

One often distinguishes the unit of assignment, the unit of intervention, the unit of observation, and the unit of analysis. The unit of assignment could be the unit that is assigned the allocation strategy, say a community is randomized to receive the vaccination strategy of interest, and another is assigned to receive a control vaccination strategy. With vaccines, the additional unit of assignment is the individual within community. The assignment at the individual level within each community may be randomized or not. For a se-

lected target group, such as children under 2 years of age, one might vaccinate whoever comes to a clinic to be vaccinated. But then the children who are vaccinated are not a random sample, but subject to a selection mechanism as in an observational study, the rule of which is unknown to the investigator. The unit of intervention could be health clinics or physician's practices within a community, or the nurse practitioner's office within a school. The unit of observation for cases is generally the individual, whether it be individual cases picked up through surveillance systems or clinical studies. However, in community studies there may also be community-level covariates, such as prevalence or incidence levels, amount of rainfall, distance from roads or distance from health clinic. So the community can be the unit of observation. There is much discussion in the literature about the appropriate unit of analysis. In general, the unit of analysis is determined by the study design. "A unit is a unit of analysis for an effect if and only if that effect is assessed against the variation among those units" (Murray 1998, page 105). Several of the design considerations in group-randomized studies to estimate the different types of effects are summarized in Tables 13.6 and 13.7.

Table 13.6. Design considerations in group-randomized studies to estimate indirect, total, or overall effects of vaccination strategies

Design consideration
Primary and secondary questions of interest
Vaccination strategy
Clinical endpoints
Study population and subpopulations
Sources of transmission
Case ascertainment
Choice of randomization unit (group)

13.3.1 Scientific or public health question of interest

Studies can be designed to evaluate direct as well as indirect, total, and overall effects. However, one of the effects may be of primary interest and another effect is or other effects are of secondary interest. For example, the primary interest may be in evaluating the total effects of vaccination compared to no vaccination, as in the pneumococcal vaccine study designed by Moulton et al (2001) (Figure 13.4). In this case, one would want to vaccinate as many individuals in the target population as possible to maximize the total effects. A similar reasoning holds if the overall effect of vaccination in the target population is of interest.

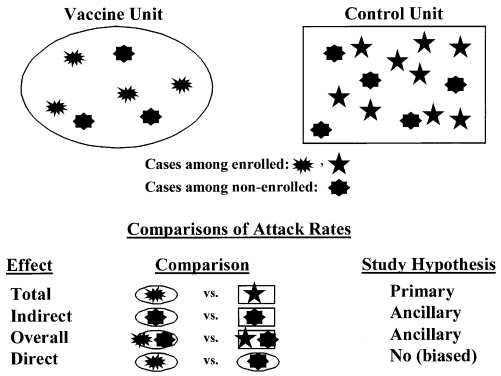

Fig. 13.4. Figure will be replaced. Schematic of the questions of interest in the pneumococcal vaccine trial in Native American. Participants in each vaccine unit receive PCV7 vaccine, and those in each control unit receive MnCC vaccine (from Moulton et al 2001).

If pure indirect effects were of primary interest, then the best approach would depend on the which subgroups were receiving the vaccines and in which subgroups the indirect effects were to be measured. For example, if they were the same subgroups, then there would be a trade-off between vaccinating too many people in the intervention communities so that there are few people left unvaccinated and few events in the unvaccinated people in the intervention community and not vaccinating enough people so that there is a measurable indirect event. On the other hand, if one were interested in estimating the indirect effects in adults of vaccinating children, then the goal would be to vaccinate as many children as possible in the intervention communities (Figure 13.5). If in addition to indirect or total effects, one is also interested in evaluating the direct effects of vaccination, then one would want to vaccinate few enough people that sufficient transmission remains to produce the number of events necessary to estimate the direct effects.

13.3.2 Vaccines and vaccination strategy

Exactly what the intervention program of interest is will depend on the vaccine, the vaccination schedule for that vaccine, and which subgroups suffer the greatest morbidity. The comparisons may be made between different levels of vaccination coverage, between allocation within different age groups or otherwise defined subgroups. In a parallel design, it might be necessary to consider using a different active vaccine as a control. This would help preserve masking and inactive placebos are often considered unethical for vaccine studies. The active vaccine as a control also provides a comparable group, in that those people who actually receive the vaccine in each group might be assumed the

appropriate groups for comparison in estimating total effects. For example, in the pneumococcal vaccine study of total effects, the control vaccine was an investigational meningococcal C conjugate vacccine (Moulton et al 2001). In a phased implementation design (Section 13.4), an active control vaccine would not be necessary.

Fig. 13.5. Possible comparisons within subgroups to estimate different effects (figure to be added).

13.3.3 Clinical endpoints

Clinical endpoints can be defined as a combination of clinical symptoms and/or by biological confirmation of the infectious agent targeted by the vaccination. The infectious agents can further be identified as being contained in the vaccine or not contained in the vaccine. For example, in pneumococcal vaccine studies, the cases can be categorized as being a vaccine serotype or a nonvaccine serotype. In influenza vaccine studies, the infections are classified either as homologous with the vaccine type or heterologous, indicating some degree of antigenic mismatch between the vaccine strains and the circulating strains.

Two possibly related problems may arise in large, group randomized studies. First, if the disease is common, such as in influenza, the number of suspected cases in the large study may be too many for all cases to be confirmed biologically. Secondly, surveillance may not be specific for the illness of interest. For instance, influenza incidence in post-licensure vaccine studies is generally measured using non-specific case definitions, such as influenza-like illness or medically-attended acute respiratory illness, which include many diseases in addition to influenza. A nonspecific case definition can attenuate the estimates of indirect and overall effects. In Chapter 8 the concept of using validation sets to obtain more accurate efficacy estimates when the main case definition is non-specific was discussed. Especially in studies to evaluate total or overall effects, validation sets might be helpful to improve the ability to detect a signal above the noise.

As an illustration, Figure 13.6 shows results of 100 stochastic simulated estimates of the indirect effects of vaccinating 50 percent of the children with an influenza vaccine in one community as compared with another community without vaccination (Halloran and Longini 2001). Each population has 10,000 people, half children and half adult. The indirect effects are set to 0.25. In each pair of populations, the population in which children were vaccinated had an influenza incidence rate reduced by a factor of 0.25. VE_S is assumed to be 0.90 (leaky). The top histograms of estimates based on ascertainment of all true influenza cases in children and adults are centered around 0.25.

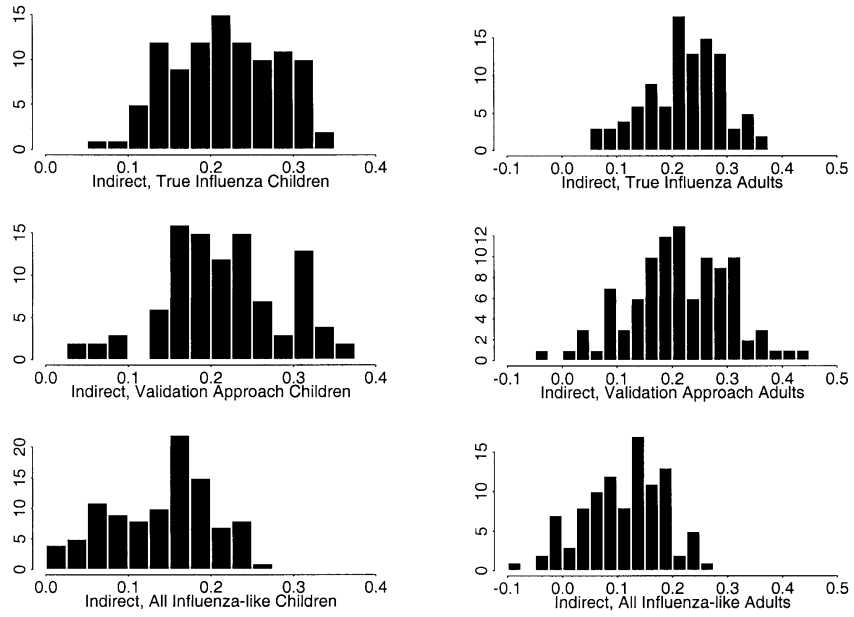

Fig. 13.6. Estimated indrect effects of vaccination of children among children (left) and among adults (right) when the indirect effects are set to 0.25. Estimates were based on true influenza cases (top), the validation set approach (middle) and all influenza-like illnesses. The expected incidence in children varied weekly over the 12-week epidemic period as (0.014, 0.024, 0.034, 0.05, 0.06, 0.055, 0.05, 0.044, 0.038, 0.024, 0.015, 0.01). The expected incidence of influenza in adults was half that. The expected incidence rate of noninfluenza in both children and adults was set to 0.02 per week. The baseline incidences of true influenza and background noninfluenza are multiplied by random numbers between 0.085 and 1.15 so the baseline incidences in each comparison pair are similar but not identical. (Halloran and Longini 2001)

However, if we use all influenza-like illnesses, the estimates are much lower (bottom rows). The histogram is centered around 0.14 in children and 0.10 in adults. However, by incorporating a random sample of the influenza-like illnesses that are biologically confirmed, we can adjust the estimates based on the influenza-like illnesses (middle row). The histograms are once again centered around 0.25, although the histograms based on the validation set approach show more variability than the histograms based on confirming all true influenza cases.

Although it may not be feasible to confirm biologically every clinically determined case in a large study to evaluate indirect, total, or overall effects, a small random sample of confirmed cases could be quite useful. There are trade-offs in using a nonspecific outcome on more groups and reducing the number of groups and using validation sets. Given the variability and background

noise between communities, schools, or other group, for a fixed budget, it is probably better to use fewer groups in general and get more specific outcomes on some of the participants. These potential trade-offs have yet to be studied rigorously.

13.3.4 Study population

The choice of study population will be determined by the vaccine and the ability to conduct a large-scale study in the population. The ability to administer the vaccine, keep records, and also to obtain data on the clinical outcomes are important.

13.3.5 Sources of transmission

Consideration of the likely transmission patterns and sources of exposure to infection in a population is required in anticipating possible detection of indirect effects. These transmission patterns will greatly influence the magnitude of the indirect effects of an intervention strategy. For example, many influenza researchers believe that school children are the primary sources of transmission in the community. Widespread vaccination of school children could be expected to have considerable indirect effects on reducing influenza in a community (Halloran and Longini 2006; Piedra et al 2007). On the other hand, in the study aimed to evaluate the total effects of vaccinating children < 2 year olds with pneumococcal vaccine (Moulton et al 2001; Moulton et al 2006), the contribution to transmission of school children or adults who are colonized with the bacteria is not well-understood. If the older children who are unvaccinated are important sources of transmission, then the vaccination strategy at least in the early years upon introduction will have low indirect effects, and the total effects will be dominated by the direct protection.

13.3.6 Case ascertainment

In large community studies to evaluate indirect, total, or overall effects, using either a pre- versus post-vacccination or a group randomized study, good methods for thorough case ascertainment is important. Examples include active population surveillance as in the Niakhar study, biological confirmation of suscepted cases in reference laboratories, or general surveillance and reporting systems. Active population surveillance can demand a lot of resources. If considerable underreporting of cases is suspected, and two or more sources of surveillance or case reporting are available, capture-recapture methods can be considered to provide better estimates of the number of cases (Gjini et al 2004).

13.3.7 Choice of randomization unit

The choice of at what level to randomize the units, that is the choice of the group, depends on both practical and theoretical considerations. One wants the groups to be transmission-dynamically separate. If communities receiving vaccine interact with communities not receiving vaccine, the contamination among groups could dilute the indirect, total or overall effects of the vaccination program. The contiguity can occur through spatial patterns or social mixing patterns among units. If there is contamination across units, then the power of the study will be diminished. The vaccination delivery system may determine the randomization units. These could be health care clinic or EPI vaccination team catchment areas. Political units such as towns or counties might be natural randomation units. One can also use smaller randomization units, such as schools (King et al 2006) or households, such as in the mini-community design (Chapter 12).

Given a study population, there is a trade-off on size of cluster with number of expected cases and the number of clusters. If incidence rates are relatively high, and the effect to be measured is also expected to be substantial, then one can divide the population into fewer clusters. However, there will be a loss of efficiency as the number of individuals per randomization unit increases. However, care should be taken that the randomization units are not too small. The efficiency of a study also depends on the intragroup correlation which could be affected by the size of the community chosen as the unit of randomization (Hayes et al 2000). If small communities are chosen, then the intracommunity correlation might be quite high, while in large communities, the correlation might be smaller. Also, small randomization units might have considerable mixing among the units, resulting in diminished indirect, total, and overall effects. In general, one would prefer to increase the number of communities to have more randomization units with fewer individuals if they are transmission-dynamically separate. The choice of the randomization unit for any particular study will depend on the local conditions.

13.4 Parallel and Stepped Wedge Designs

The three general approaches to designing group-randomized studies are parallel designs, stepped wedge designs, and cross-over designs (Table 13.7)(Hughes 2003). Any of these three study types can be used when the randomization unit is either an individual or a group, but the focus here is the context of group-randomized studies to evaluate indirect, total, and overall effects. In the parallel design, the groups are randomized to receive one or other of the interventions at the beginning, and the intervention assignment does not change until the end of the trial. In the stepped wedge design, the intervention is introduced in more and more groups over time. This allows the groups in which the intervention is not yet introduced to serve as control groups. In the

crossover design, the groups are first randomized to receive one or other of the interventions at the beginning, then at some point, the interventions are switched. This latter design likely has no application in vaccine studies, since in general one cannot de-vaccinated people or populations. We do not consider it further here. Both the parallel design and the stepped wedge design can be used to evaluate direct, indirect, total, and or overall effects, depending on the design and implementation. That is, groups can be randomized to vaccination or control, then individuals within groups may or may not be randomized to receive vaccine or not.

Table 13.7. Community-randomized designs and randomization schemes

Design	Randomization scheme	Covariate constraints
Parallel	Completely randomized	Unconstrained
Stepped wedge	Stratified	Constrained
(Cross-over)	Matched-pairs	

13.4.1 Parallel Designs

The simplest parallel group randomized design is one in which N groups are randomized to either vaccination or control, for a total of $2N$ groups. It would also be possible to have an unbalanced allocation.

13.4.2 Parallel Pneumoccocal Vaccine Study

Moulton et al (2001) designed a group-randomized, double-masked phase III trial of a *Streptococcus pneumoniae* conjugate vaccine in American Indian populations in the US. The study had a parallel design. The goal of the trial was to evaluate the total effects of vaccination as well as the indirect effects, and at the same time to serve as a pivotal vaccine study. At the time of the design of the study, another phase III study with standard individual randomization was ongoing in northern California (Black et al 2000). However, the number of invasive pneumococcal cases occurring in that trial was small. The group randomized study was designed to estimate the total efficacy, which takes the direct protective effects on the vaccinated individuals as well as the indirect effects into account, so that the effects could potentially be greater than the effects in the individually-randomized study. The study was the first group-randomized vaccine trial in the United States designed to be a pivotal trial for licensure.

There were 4,164 infants enrolled in the PCV7 communities and 3,926 in the MnCC communities between April 1997 and December 1999. The study had 38 geographically defined randomization units which had been formed

to minimize the degree of social mixing, and hence contamination, between randomization units. Half the units were randomized to study vaccine, a seven-valent conjugate pneumococcal vaccine (PCV7 vaccine), the other half to an active control, a conjugate meningococcal group C vaccine (MnCC vaccine). The goal in each randomization unit was to vaccinate as many children under 2 years of age as possible to achieve the highest total effects.

Originally the trial was designed to continue until 48 cases of invasive pneumococcal disease due to vaccine serotypes had accumulated. However, on February 17, 2000, the FDA approved the licensure of the PCV7 vaccine based on the results of the primary efficacy study in northern California (Black et al 2000). Ethically the study could not be continued and PCV7 vaccine was offered in the MnCC communities. Only nine cases had accrued at that time.

Later, Moulton et al (2006) estimated the indirect effects on the unvaccinated children in the communities (Section 13.7.2). To estimate indirect effects they compared the incidence rate of invasive pneumococcal disease in vaccine units among non-enrolled children versus the incidence rate in control units in non-enrolled children. By using the non-enrolled children in both communities, they hoped to have comparable children in their analysis. By combining the information from the study with information from Indian Health Service User Population data and birth logs, they were able to obtain denominators for each of the 38 randomization units. They were also able to interpolate the number of non-enrolled children at any day btween April 1997 and October 2000. The numerator for invasive disease was obtained from surveillance data that had been subject to a standard protocol during the study. There were 21 cases of invasive disease due to study vaccine serotypes among nonstudy children living in MnCC randomization units, and 27 cases among those in the PCV7 units.

13.4.3 Stepped Wedge Designs

Stepped wedge designs can be used when a parallel design is unfeasible either for practical or for ethical reasons. For example, if a vaccine is already licensed, then it may be unethical to randomize some communities or individuals not to receive vaccine during the trial. It may be that practical considerations preclude introducing the vaccine everywhere at once, either because insufficient vaccine is available or for logistic reasons of not being able to administer it everywhere or to everyone at once. By the end of a trial using a stepped wedge design, all randomization units will have received the vaccine. Thus the clusters are not randomized to receive the vaccine intervention or not, but rather the time of the introduction of the vaccine intervention to each cluster is randomized (Figure 13.7).

Considerations of design and power in the stepped wedge design revolve around the timing of the individual observations, the interval at which the intervention is introduced into groups, and the number of groups switched from control to intervention at any given time. Observation of individuals

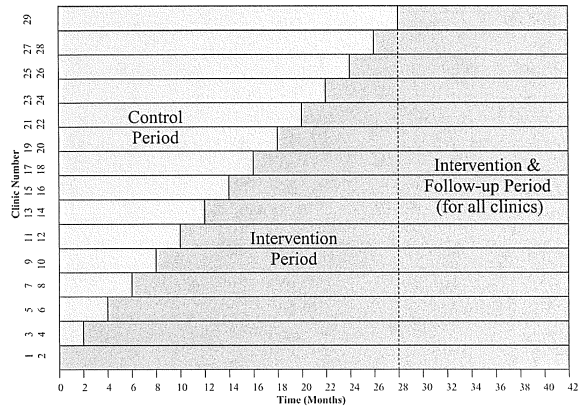

Fig. 13.7. Example of a stepped wedge design (this figure will be replaced with an original figure. from Moulton et al 2007)

could occur continuously, or somehow be aligned with the timing of switching the groups from control to intervention. The number of time points chosen to introduce a given number of clusters into a trial influences the power of the study. The higher the number of time points, the higher the power, especially if the number of observations on individuals is correlated with the number of time points (Hussey and Hughes 2007). However, if individuals are observed continuously, then there is less effect on power. In vaccine studies in which cases are reported as they occur, the effect on power would be lower.

The stepped wedge design is gaining in popularity (Moulton 2007; Hussey and Hughes 2007; Hughes 2008), though at the time of this writing, we do not know of any group randomized vaccine trial in which this design was used to estimate indirect, total, or overall effects. Not all outcomes of interest can be studied by a stepped wedge design. For example, if one is interested in the change in age of first infection, then one would possibly conduct the study over several years. If a vaccine is unlicensed, this may be unfeasible, since one would not want to wait for years to license the vaccine on this outcome. If a vaccine is licensed, such a long term study would likely not be ethical. Thus, the observational studies of pre- and post-vaccination will have to suffice (Section 13.2.1) or mathematical models can be substituted as a means of experiment (Chapters 4 and 5).

13.4.4 The Gambia Hepatitis Intervention Study

One of the first stepped wedge designs was a hepatitis B vaccine study in the Gambia (Gambia Hepatitis Study Group 1987). Though this study was not designed to evaluate indirect or overall effects of vaccination, it is presented in

this chapter because of its early use of the stepped wedge group-randomized design.

Chronic liver disease and liver cancer, are thought to be partially caused by hepatitis B viral infection. In West Africa, including The Gambia, chronic liver disease and liver cancer are important public health problems. It used to be that nearly everyone in The Gambia was infected with HBV during childhood and between 10 to 20% became chronic carriers. The goal of the hepatitis B vaccination study was to evaluate the effect of infant vaccination on preventing chronic liver disease and liver cancer later in life. Thus, a long-term follow-up for over 30 years was planned. However, it was undesirable to do a parallel randomized study in which half of the children were followed for 30 to 40 years before initiating mass vaccination campaigns. Thus, a phased implementation, or stepped wedge design was proposed. Because at that time, four injections were required for full immunization, and the vaccine was to be administered along with the routine EPI vaccines, it was considered logistically unfeasible to do an individually randomized trial, as well as potentially ethically questionable.

The choice of study designs was further influenced by the expense of the vaccine and its limited availability prohibiting immediate universal hepatitis B vaccination. To avoid confounding by secular trends, the stepped wedge design provided the ability to have comparison groups available from the same time period. They also hoped that the hepatitis B vaccine would be widely available by the end of the study. Based on these considerations, phased introduction of hepatitis B vaccine to the EPI schedule was planned, with injections within one month of birth, and at 2, 4 and 9 months of age. There were 17 EPI vaccination teams each assigned a portion of 104 delivery points that were visited at least once every two weeks. The study plan randomized one of the teams every 10 to 12 weeks to introduce the hepatitis B vaccine to the EPI schedule by vaccinating all newborns who report to the vaccination points served by the team. This was to continue for a period of about four years, when all teams would be giving the vaccine, so that country-wide coverage would be achieved (Figure 13.8). The alternative parallel design, in which EPI vaccination teams would have been randomized to give HBV or not for four years is statistically more powerful, but would have been less acceptable (Jaffar et al 1999).

Evaluation of the protective effect of HBV vaccination against liver cancer and chronic disease was planned through the long-term follow-up of those children born during the 4-year period over which HBV vaccine was introduced. For children born in each three month period, incidence of later liver cancer and chronic liver disease would be compared among those receiving HBV vaccine and those not. For example, those newborns entering in the first three months of the study would be compared later in life to those newborns reporting to the 16 other vaccination teams. This approach to comparison controls for secular trends that might affect the risk of developing liver cancer. Randomization of the order in which the EPI teams introduce the HBV

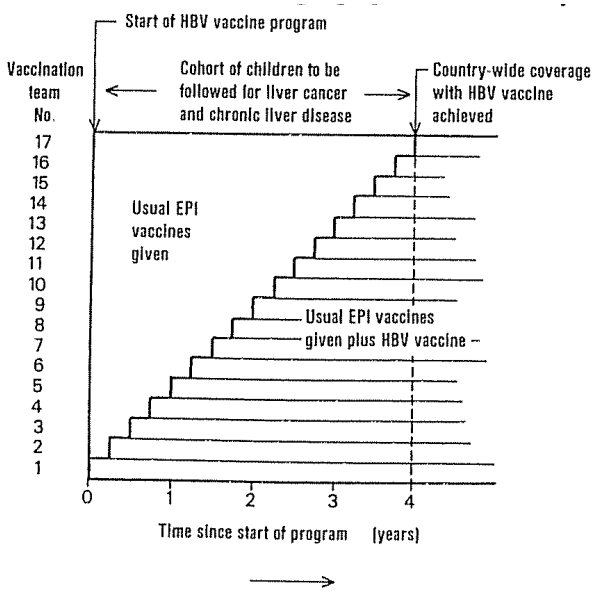

Fig. 13.8. Stepped wedge design in the Gambia (from the Gambia Hepatitis Study Group 1987)

vaccine minimizes the bias in the comparison of the vaccinated and unvaccinated groups. To further avoid bias, the plan is to restrict the analysis to comparison of those who attend the vaccination clinics at all four ages at which HBV would be given.

Considerable efforts were undertaken to enable identification of the persons enrolled 30 to 40 years after enrollment, which in The Gambia can be a challenge. These efforts make interesting reading but are not relevant here. For follow-up, a nationwide cancer registry and active surveillance were established. A number of studies to assess intermediate endpoints were built into the long-term follow-up. A subset was followed for serological data on a regular basis through childhood and adolescence. Cross-sectional studies were also performed to compare acquisition of HBV markers at different ages. This is a superb example of the need to plan for long-term studies and follow-up in vaccine studies.

13.5 Randomization

13.5.1 Approaches

Group-randomized trials often have only a limited number of identifiable groups to assign to the different interventions. Two key issues arise in choosing a randomization scheme when the number of groups is limited. First, in

a completely randomized study, variability among communities could swamp out the estimates of the effects of the vaccination strategy. In a completely randomized parallel study, groups are randomized to intervention or control without any consideration of variability among the groups. In a completely randomized stepped wedge design, the order of introduction of the intervention is randomized without any consideration of variability among the groups. Secondly, generally there will not be enough groups to ensure that the potential sources of bias among the intervention conditions will be evenly distributed. Even if the groups or communities contain thousands of participants, if there is important variability of characteristics between groups, a study that is completely randomized at the group-level could have imbalances in important covariates. If these characteristics are also related to the outcome of the study, an example would be the incidence rate of the disease of interest in the community, then the results of the study may be difficult to interpret. Even if it is possible to do some adjustment at the time of analysis, the results will be open to criticism. For example, consider a study to evaluate the total effectiveness of widespread vaccination, with five communities randomized to intervention and five to control. It could happen that the five communities with the lowest baseline incidence of the disease in question would be randomized to receive vaccination. If that were to happen, the results of the study could be criticized as being biased in favor of the vaccination strategy by the realized randomization.

Two main approaches have been proposed to increase power and to reduce the chance of unbalance on covariates when there is considerable variability among communities. One is to stratify groups by pre-randomization group-level covariates of interest, including transmission characteristics, then randomize to intervention or control within strata. The other, an extreme version of stratification, is to match pairs of communities on the covariates of interest, so the strata contain only two communities, then randomize to intervention or control within the pairs. Groups rather than individuals are stratified/matched prior to randomization. Hayes et al (1995) matched on transmission characteristics in a community trial of the effect of improved sexually transmitted disease treatment on the HIV epidemic in rural Tanzania (Grosskurth et al 1995). Pre-randomization stratification or matching requires information on factors related to the primary endpoint used for the stratification/ matching prior to randomization.

13.5.2 Covariate-constrained randomization

An appealing approach to randomizing groups that avoids gross imbalances on known and measured variables is covariate-based constrained randomization (Moulton 2004). In the constrained, or restricted randomization, certain balancing criteria are determined before randomization that still retain validity of the design. Then the final randomization scheme is randomly chosen from among those that both satisfy the constraining criteria and are still valid. A

completely randomized design is valid if each pair of randomization units has the same probability of being allocated the same treatment (Bailey 1983). A design is biased, if, across the randomization units, there is any difference in probability of assignment to a given treatment. Even in stratified or pair-matched designs, an unlucky randomization can result in the intervention always being assigned to the lower incidence groups. Constrained randomization can be used in completely randomized, stratified, or pair-matched designs. We consider it primarily in the context of a completetly randomized parallel design. We then consider briefly using constrained randomization in the stepped wedge design.

Different constraints can be used for different types of constraining variables. For continous covariates such as incidence rates of the disease of interest, one can choose some measure based on the standard deviation or absolute mean difference. For dichotomous covariates, $\pm$ some percentage points might be appropriate. For example, suppose there was a difference in the incidence of disease between the north and south regions of the study area. Then one would not want all of the intervention sites in the north and control sites in the south. One could assign a 0,1 dummy variable for north and south and require that the difference between the intervention and control values be less than 10%. Other important aspects, such as sources of water, proportion of the population with a certain educational level, health clinics, or roads within geographic areas can also be balanced within some specified range. Composite scores or more than one covariate can be used for defining the constraints that need to be satisfied. The constraining criteria can vary among the covariates.

Once constraints are set, then one needs to identify all of the possible allocations that satisfy the constraints. To do this, form a list of all the possible allocations. For a completely randomized (at the group level) design, there will be $\binom{2m}{m}$ entries, where $2m$ is the total number of groups. For a pair-matched design, there will be 2^m entries, where m is the number of pairs. Making a pass through all of these entries, select those allocations that meet the specified criteria. These criteria could mean achieving some level of balance on a given set of covariates.

Once the allocations that meet the set of constraints have been identified, they need to be checked to see whether the possible allocations meet the requirement of validity of the randomization scheme. For example, some pairs of groups may always be in the same arm of the study, while others may never be in the same arm. To check the allocations, make a matrix whose elements are the number of times, from among those allocations satisifying the constraints, each pair is together. Examine the list for signs of over- or underrepresented pairs. If the allocations seem overly constrained, then relax one or more of the constraining criteria. Identify the allocations that satisfy the new constraints, and check them once again. Repeat the relaxation of the constraints until the allowable allocations seem appropriate. Then randomly select one of the allowable allocations. If there are too many possible allocations to enumerate, one can construct the matrix from a large number of acceptable designs, and

choose one of them. A SAS program is available to perform this algorithm (Chaudhary and Moulton 2006).

13.5.3 Hypothetical dengue vaccine study

As a simple example, suppose that we are designing a dengue vaccine study in four communities where interest is in the overall effects of vaccination. Two of the villages will be randomized to receive vaccine and two not. This small number of communities is chosen only for illustrative purposes. Generally more communities would be required. The expected annual incidence of dengue in each community is correlated with the outcome of interest, thus there is concern about the balance of the dengue incidence in the two vaccine communities and the two control communities. With four communities, there are six possible unique allocations of vaccination and control (Table 13.8). Baseline surveillance over the three previous years yielded estimates of average annual incidence 3, 5, 11, 14% in the four communities. If no constraints were placed on the randomization, then one of the six allocations would result. However, in allocation A, the two communities with the lowest incidences receive vaccine, and in allocation B, the two communities with the highest incidences receive vaccine. The mean absolute difference in baseline incidence is 8%, higher than the overall average incidence. There is a 1 in 3 chance of selecting one of these randomizations. Alternatively, one could say that only those allocations are acceptable that yield exact balance on the average annual incidence. In this example, allocations C and D satisfy this constraint, though generally one could not expect that any allocation would yield an exact balance.

However, the problem now is that in allocations C and D, the two communities with incidence rates of 3% and 13% and the two communities with 5% and 11% are always together. This violates the validity principle stated above, since for example the pair 5 and 13 do not have a chance of being randomized together. In essence, each pair of communities in allocation C and D is acting as a single community. To alleviate this problem, the constraint could be relaxed, so that the mean difference in annual incidence is less than 3 percent. Then allocations B, C, D, and E would satisfy the contraint. Though the communities with 3% and 5% and those with 11% and 13% could never be together, this is the same as would happen if it were a pair-matched design, (3% and 5%, 11% and 13%) with randomization within pairs. More details are in Moulton (2004).

13.5.4 Constrained randomization for a stepped wedge design

One would also like to achieve balance in group-level covariates when randomizing the sequence of groups converting from control to intervention in stepped wedge design. For example, it would be undesirable for all of the low-incidence communities to be randomized to introduce the vaccination strategy early in the stepped wedge study. One might want to aim for a balance on

Table 13.8. Baseline average annual dengue incidence rate (percent) over the past three years in each of six communities to be included in the dengue vaccine trial. The balance of the randomization is measured by the mean difference in average annual incidence between the communities to receive vaccine and the control communities. (adapted from Moulton (2004))

	Communities				
Allocation	Vaccine		Control		Mean difference
A	3	5	11	13	−8
B	3	11	5	13	−2
C	3	13	5	11	0
D	5	11	3	13	0
E	5	13	3	11	2
F	11	13	3	5	8

group-time spent in the control and vaccination program status with respect to the group-level covariates of interest. Moulton et al (2007) developed a method for constrained randomization in the stepped wedge design of a study introducing screening for tuberculosis in HIV clinics in Rio de Janeiro, where more details are found. The general idea in designing the stepped wedge constrained randomization is that for each possible sequences of introduction of the vaccination strategies, the constraints are checked to see whether they are satisfied. If the number of groups is too large to enumerate all possible sequences, then sequences are sampled randomly from all possible ones by random permutations of the group labels. For each permutation, the constraints are checked to see whether they are satisfied. Then when a large number of acceptable sequences have been identified, one is randomly selected from it.

Moulton et al (2007) suggested the following ad hoc approach to checking the constraints. For each jth covariate of the ith group, $i = 1, \ldots, N$, x_{ij}, for a given time of entry t_j, $t = 1, \ldots, T$, let c_j be a proportional covariate-specific tolerance. The constraint can be expressed as

$$\frac{1}{1+c_j} < \frac{\sum_{i=1;t_i \neq T}^{N}(T-1-(t_i-1))x_{ij}}{\sum_{i=1;t_i \neq T}^{N}(t_i-1)x_{ij}} < (1+c_j). \tag{13.1}$$

Then the sum of the covariate values weighted by the number of time units in the intervention status must be within $c_j \times 100\%$ of that for control status. A similar approach to that described in Section 13.5.2 is followed. One tries to avoid constraints that always pair two groups to enter simultaneously, as this would effectively reduce them to a single randomization unit.

13.6 Power and Number of Communities

13.6.1 General considerations

In group-randomized studies, the sample size calculation needs to take into account that randomization is by group rather than by individual. In general, group-randomized designs are less efficient than individually randomized studies due to the related factors of intra-group correlation and intergroup variability. That is, the more similar the individuals within each group are to each other and the more different the groups are from one another, the greater the group design effect on sample size will be. For a given sample size, a stepped wedge design will generally be less efficient than a parallel design, so further allowance needs to be made when planning a stepped wedge design study.

Two different measures are used in calculating sample size for group randomized studies. One is the coefficient of variation k, the standard deviation/mean of the incidence rate, or other outcome measure of interest such as proportions (attack rates) or mean of a continuous variable in the groups in the study. Another approach uses the design effect D, or variance inflation factor, σ. For trials with equal numbers of individuals in each community,

$$D = \sigma = 1 + (n-1)\rho, \tag{13.2}$$

where n is the number of individuals per community, ρ is the intra-cluster correlation coefficient, and D is the factor by which the sample size needs to be increased above that required for an individually randomized trial to make up for randomization by cluster (Donner and Klar 1994).

We consider sample size calculations based on incidence rates, proportions (attack rates), and means of continuous outcomes. For clarity, the following discussion is just about rates, but could applies to proportions (attack rates) and means as well. The sample size calculations require estimates or assumptions about the baseline incidence rate λ_0, and an assumption of the effect of the intervention strategy, or equivalently, the rate in the intervention group λ_1. Exactly what the λ_0 and λ_1 of interest are will depend on whether the primary interest is on estimating indirect, total, or overall effects, or possibly even direct effects. For example, if the total effect of a vaccination strategy is of interest, then λ_0 might be the incidence rate in the children receiving a control vaccine, and λ_1 the incidence rate in children receiving the vaccine of interest. If overall effects were of interest, λ_0 and λ_1 could be the incidence rates in all age-appropriate children (or all children) in the control and the vaccination intervention groups. If the indirect effects of vaccinating schoolchildren against influenza on the incidence rates in adults were of interest, then λ_0 and λ_1 could be the rates in the adults in the control and intervention groups. If more than one effect is of interest, then sample size calculations can be made for more than one effect.

Hayes and Bennett (1999) provide simple formulae to determine sample size for parallel design group-randomized studies. The next section is based primarily on their paper, where further details, references, and examples are available. Many group-randomized studies of vaccination strategies may require more complex computations than these. In some cases, stochastic simulations of the populations with the planned intervention strategies can be used to estimate expected effects, power, and sample sizes for the studies (Halloran, et al 2002). Sample size requirements under randomization tests are similar as those for model-based inference procedures (Murray 1998, page 117).

13.6.2 Sample size calculations for parallel unmatched studies

Assuming that there are equal numbers of groups in the intervention and the control arm, let N be the number of groups in each study arm. Then the total number of groups in the study is $2N$. Let $z_{P=2}$ and z_β be the standard normal distribution values corresponding to upper tail probabilities of $\alpha/2$ and β. The corresponding sample size will give a power of $100(1-\beta)\%$ of obtaining a significant difference ($P < \alpha$ on a two-sided test), assuming that the true population rates in the intervention and control groups are λ_1 and λ_0. If the outcome is based on person-time, let y denote the person-time of follow-up in each group. Then the number of groups required in each arm is

$$N = 1 + (z_{P=2} + z_\beta)^2 \frac{(\lambda_0 + \lambda_1)/y + k^2(\lambda_0^2 + \lambda_1^2)}{(\lambda_0 - \lambda_1)^2}. \tag{13.3}$$

If the outcome is based on proportions (attack rates), let π_0 and π_1 be the true population proportions (attack rates) in the intervention and control groups. Let n be the number of individuals in each group. Then the number of groups required in each arm is

$$N = 1 + (z_{P=2} + z_\beta)^2 \frac{\pi_0(1-\pi_0)/n + \pi_1(1-\pi_1)/n + k^2(\pi_0^2 + \pi_1^2)}{(\pi_0 - \pi_1)^2}. \tag{13.4}$$

If the outcome is based on a continuous reponse, such as parasite density, then the objective is to compare the mean of that variable in the vaccine intervention and control groups. Let μ_1 and μ_0 be the true population means and σ_1 and σ_0 be the within-group standard deviations of the outcome variable in the intervention and control groups. Let n be the number of individuals in each group. Then the number of groups required in each arm is

$$N = 1 + (z_{P=2} + z_\beta)^2 \frac{(\sigma_0^2 + \sigma_1^2)/y + k^2(\mu_0^2 + \mu_1^2)}{(\mu_0 - \mu_1)^2}. \tag{13.5}$$

If one is interested in direct protective effects, these equations are analogous to those for individually-randomized trials in equations 6.12, 6.13, and 6.14. The design effect associated with the group-randomization can be estimated by dividing the equation in this chapter by the corresponding equation in Chapter 8.

13.6.3 Sample size formulae for parallel pair-matched studies

When pairs of groups are matched before randomization on the basis of factors expected to be correlated with the main study outcomes, the hope is the matching will minimize the degree of between-group variation within matched pairs. However, there is a trade-off between the increase in power and precision by increasing the comparability of the intervention and control groups and the loss of power due to the reduced degrees of freedom that is well-discussed in the literature (Martin et al 1993; Hayes et al 1995). Much has been done on characteristics of the general size and correlation between the endpoint of interest and matching covariates and power in cluster-randomized trials in general (see for instance Murray 1998). Hughes (2005) more generally considers using baseline data in designing a group randomized trial to choose between an unmatched or pair-matched design, choice of effect measure, and the power to be expected from the various strategies. Equations 13.3, 13.4, and 13.5 can be adjusted to take account of matching with two changes. First, to adjust for the required number of degrees of freedom, add 2 instead of 1 to required number of groups in each arm (Snedecor and Cochran 1967). Secondly, the coefficient of variation k is replaced by k_m, the coefficient of variation in true rates (or means or proportions) between groups within the matched pairs prior to intervention.

13.6.4 Coefficient of variation

Since a value for the coefficient of variation is needed for the sample size calculations, in the absence of any empirical data, an assumption about the value must be made. In this case, one can compute power curves and examine the number of clusters required for plausible values of k. Sometimes data may be available from baseline surveillance studies. Alternatively, data may be available from a pilot study conducted to check the implementation plan that is also used to collect data to estimate the inter-group variability of the main outcome of the trial. A subset of the groups can be selected and data on a small fraction of the population of interest be recorded. Alternatively, data might be available on similar groups in different areas of the country. Hayes and Bennet (1999) provide formulae for estimating the coefficient of variation for unmatched (k) and matched (k_m) studies. Generally k will be larger than k_m. The coefficient of variation is for the variation in the true rates between groups, not the variation in the estimated rates which contains an element of within-group random variation. The general idea is to compute the empirical variance of the group-specific results, then substract the component of the variance due to sampling error. Moulton et al (2007) has an example.

13.6.5 Another approach

In another approach to sample size calculation for a group randomized study, one might compute the number of events needed under individual random-

ization to achieve a certain power, possibly for the lower bound of a 95% confidence interval to lie about a certain pre-determined efficacy if, in fact, the true efficacy is some other higher efficacy. Then, to account for intra-group correlation, multiply the number of events by the usual design effect in equation (13.2). One can possibly get an initial estimate of the overdispersion σ^2 directly from some baseline data from a sample of the communities (Moulton et al 2001).

13.6.6 Sample size for stepped wedge design

To take account of the stepped wedge design in the sample size, Moulton et al (2007) suggest a modification to equations (13.3)–(13.5). Essentially the standard deviates $z_{P=2}$ and z_b used in equations (13.3)–(13.5) are multiplied by a factor > 1 that accounts for the lower efficiency of the stepped wedge design. In addition, if the variability between groups is large, they suggest substituting the harmonic mean for the simple mean.

The multiplicative factor can be computed in various ways. As an example, consider a stepped wedge design study in which the main analysis was to compare the incidence in groups receiving the vaccination intervention to those not yet receiving the vaccination intervention based on a conditional likelihood as in equation (13.9). Following Moulton et al (2007), one could use a log-rank test statistic, which is essentially the score test statistic for such a model with one treatment covariate. One weights the hazard function within the time unit of interest (week, month) by the log-rank to estimate the effect of the stepped wedge design. Let T be the last time unit at which control groups begin the intervention. Let d_{T_i} be the number of incident cases in the ith time unit in the intervention groups and Y_{T_i} be the number of persons at risk in those groups, and d_i and Y_i be the cases and persons in both intervention and control groups in the ith time unit. Then the log-rank test statistic is

$$Z = \frac{\sum_{i=1}^{T-1}[d_{T_i} - Y_{T_i}(d_i/Y_i)]}{\{\sum_{i=1}^{T-1}(Y_{T_i}/Y_i)\,(1-(Y_{T_i}/Y_i))\,(Y_i-d_i)/(Y_i-1)\}^{1-2}d_i} \qquad (13.6)$$

The statistic (13.6) can be computed by generating data sets under two different assumptions. First assume that the number of persons at risk in intervention and control groups are equal and constant over the course of the study, simulating a time-uniform equal allocation parallel design study, and yielding Z_E. Second, generate data so that the persons at risk in each month in the intervention community increases in each time unit according to the plan of the phased implementation to yield Z_{SW}. In general, for such hypothetical studies, given the same sample size, incidence and effectiveness, stepped wedge study's test statistic will be smaller than that for a parallel study by a factor of Z_{SW}/Z_E, where Z_{SW} is always smaller than Z_E (Moulton et al 2007). To account for a stepped wedge allocation, multiply the standard normal deviates in equation (13.3) by a factor of Z_E/Z_{SW}. Finally, one should

vary the values of the coefficient of variation, the assumed incidence rates, and the assumed effectiveness of the intervention to determine the range in which one would have the desired power and Type I error, then examine whether these conditions are feasible under the conditions of the proposed study.

13.7 Analysis

13.7.1 General considerations

The key issue in analyzing group-randomized studies is to account for the clustering or group-randomization. The variability of the estimates is determined not only by the number of individuals in the study, but the amount of intra- and intergroup variability. There are two general approaches to analysis that account for potential within-cluster correlation (Donner et al 1994). One approach is to reduce the data for each cluster to a single observation and to perform a standard two-sample analysis. Another approach is to do the analysis at the individual level but account for correlation somehow. Correlation within the units could be taken into account by doing a bootstrap (Efron and Tibshirani 1993) at the level of the entire community (Halloran et al 2003; Moulton et al 2006) (see Section 12.2.1). One could fit a random effects model or used generalized estimating equations. Another approach is to use a robust variance estimator (Moulton et al 2006).

The stepped wedge design trials present additional complications. Each randomization unit spends time in both the control and intervention conditions. There could be substantial secular trends in the incidence of the disease of interest, confounding the treatment effect. Moulton et al (2006) take an approach that compares the outcomes at any point in time across all groups, then combines the results over time at the same time accounting for within-cluster correlation (see below). They accomplish this by conditioning on each time unit of the study and comparing incidences in those groups that have not introduced the intervention with those that have. The analysis is carried out by maximizing a partial likelihood function that is similar to a Cox proportional hazards model.

13.7.2 Pneumococcal vaccine study

One approach is to use a model based on a non-homogeneous Poisson process in time and space (Moulton et al 2006). Let λ_{it} be the rate of disease among the individuals of interest in randomization unit i on day t. A simple model for λ_{it} is given by

$$\lambda_{it} = n_{it} \exp(\alpha_t + \gamma z_i), \tag{13.7}$$

where n_{it} is the person-days of exposure in the ith group on the tth day, α_t represents the effect of the tth day, and γ is the log rate ratio comparing

those in the intervention communities ($z_i = 1$) to those in the control unit ($z_i = 0$). The parameter α_t is a nuisance parameter that captures any day-specific secular trends, such as seasonal or weekend effects.

If living in the intervention community (and possibly also receiving the vaccine) confers protection on the individuals of interest, then γ will be negative. One can imagine a number of different comparisons, depending on whether one is trying to estimate indirect, total, or overall effects. Moulton et al (2006) were interested in estimating the protective indirect effects on invasive disease for non-enrolled children under 2 years of age.

The problem with model (13.7) is that it does not allow for different levels of coverage among the randomization units. One option is to group the coverage levels or enrollment levels, and to use dummy variables in the model that are crossed with the dummy variable for treatment arm. Moulton et al (2006) fit the model

$$\begin{aligned}\lambda_{it} = n_{it}\exp(&\alpha_t + \beta_1 Mnc_{it}^{25\ \ 49} + \beta_2 Mnc_{it}^{50+} \\ &+\beta_3 Pnc_{it}^{0\ \ 24} + \beta_4 Pnc_{it}^{25\ \ 49} + \beta_5 Pnc_{it}^{50+})\end{aligned} \tag{13.8}$$

where $Mnc_{it}^{25\ \ 49}$ is the unity for the ith unit on the tth day if it is a community randomized to MnCC vaccine, and if 25–49% of the children under age 2 on that day have received at least one immunization, otherwise it is zero. Since the communities were not randomized to different coverage levels, there may be unmeasured confounders associated with the coverage levels. So then it is of particular interest to compare across treatment arms within coverage levels. For example, if the difference $\beta_4 - \beta_1$ is negative, then it suggests presence of indirect effects at that level of coverage. The rate ratio comparing the two treatment arms at above 50% coverage is given by $\exp(\beta_5 - \beta_2)$.

Moulton et al (2006) suggest an analytic strategy that eliminates the nuisance parameter α_t by conditioning on each day of the study. The approach is similar to that in a Cox regression model where each day delineates a risk set, similar to a stratum in a case-control study. The characteristics of those randomization units that experienced a case on that day are compared to those that did not have any cases on that day. This is done for each day, and then the probabilities are multiplied together to get the conditional likelihood function:

$$\prod_{t=1}^{t=T}\left[n_{it}\exp(\boldsymbol{x}_{it}\boldsymbol{\beta}) / \sum_{j\,2\,R(t)} n_{jt}\exp(\boldsymbol{x}_{jt}\boldsymbol{\beta})\right]^{\check{Z}_t} \tag{13.9}$$

where T is the number of days in the study, δ_t is one if there is a case on the tth day and zero otherwise, $R(t)$ is the set of indices of those units at ridk on day t, and $\boldsymbol{x}_{jt}$ is the row vector of summy variables for the jth unit on day t, with $j = i$ representing the community with a case on that day. The conditional likelihood function is maximized to get estimates of $\boldsymbol{\beta}$. The computation can

be done using software for conditional logistic regression with an offset term of $\ln(n_{it})$.

Table 13.9 contains the results of fitting the conditional logistic model wtih linear predictor as in equation 13.8. The analysis did not yield significant indirect effects on non-enrolled children. One can compare the units with similar coverage levels. For example, at the coverage levels >50% , $\exp(\hat{\beta}_5 - \hat{\beta}_2) = \exp(1.96 - 193) = 1.03$. Using the naive covariance matrix for the parameter estimates yields a 95% Wald interval for the ratio 1.03 or (0.31, 3.45). The issue may be that the proportion of the population vaccinated was quite small, and that carriage from older siblings could have been important.

Table 13.9. Analysis results from fitting conditional logistic models with five dummy variables to represent six vaccine arm/percentage vaccine coverage combinations. Conditional maximum likelihood estimates, standard errors, and 95% bootstrap percentile intervals. The reference category are units that received MnCC vaccine which on a given day had less than 25% of children enrolled in the study. The CMLEs are the log rate ratios comparing incidence in non-enrolled childrein in the given category in the reference category (from Moulton et al (2006)).

Dummy variable (arm/% coverage)	CMLE	Naive SE	Bootstrap SE	Naive CI	Bootstrap percentile interval
MnCC 0–24%	0*				
MnCC 25–49%	1.18	0.64	0.62	−0.08, 2.43	0.12, 2.74
MnCC 50+%	1.93	0.83	0.81	0.30, 3.56	0.46, 4.25
PCV7 0–24%	1.09	0.59	0.60	−0.06, 2.24	−0.07, 2.58
PCV7 25–49%	0.98	0.66	0.75	−0.32, 2.28	−1.05, 2.59
PCV7 50+%	1.96	0.75	0.85	0.50, 3.43	0.68, 4.37

13.7.3 Other approaches

Ali et al (2005) entered coverage level as a continuous variable in the cholera study. In a community randomized study, one would also add a variable for treatment arm. One could deal with a secular trend by examining rate changes for groups and months when the treatment status is the same, then adjust for the estimated trend. This approach might produce results that are difficult to interpret if there is no smooth trend. The model of α_t assumes that the secular trends represented by α_t are the same for all randomization units. This might not be the case if a study such as for a meningococcal vaccine were being done on different continents. However, then a more complex model that allowed for some continent or geographic specific secular trends might be possible. One might also consider a combination of matched-pair design and analysis, even in the case of a stepped wedge design. Certain other aspects, for example,

that immunization might not begin simultaneously in all units can be taken into account by entering the randomization unit into the analysis on the day of the first immunization in the unit.

13.8 Causal inference for indirect, total, and overall effects

13.8.1 General approach

In Section 13.1.1 we informally defined direct, indirect, total and overall effects using concepts from the potential outcome approach to causal inference. In Chapters 9 and 15 we use causal inference to define estimands of interest. Defining causal estimands for indirect, total, and overall effects using potential outcomes is not straightforward. The approach assumes that individuals could potentially receive each of the treatments under study and that each of those treatments could be enumerated. Generally the assumption is made that the outcome in one individual is independent of the treatment assignment in the other individuals in the study population. This is called the assumption of no interference (Cox 1958) and is an essential aspect of the stable unit treatment value assumption (SUTVA) (Rubin 1978). Under the assumption of no interference, if there are two treatments, such as vaccine and control, then a person has two potential outcomes, one for each treatment.

The general approach in causal inference using potential outcomes is to define causal estimands and the conditions under which they can be identified from the data. One has a population of individuals. The individual causal effect can be defined, but it is not identifiable. An average causal effect estimand for the population is defined that is also not identifiable. Under the assumption of no interference and a posited assignment mechanism, such as randomization of individuals to either treatment, then the average causal effect in the population is estimable from the observed outcomes.

In the dependent happenings in infectious diseases, the assumption of no interference does not hold and indeed is the source of the indirect, total, and overall effects of interest in this chapter. The vaccine status of other individuals in the population can affect the potential outcomes of an individual, so a person can have many more than two potential outcomes, depending on the vaccine assignment to the other individuals. Rubin (1990) suggested a general notation in which the potential outcome of a person was defined as a function of the vector of treatment assignment to the person of interest as well as the treatment assignments to other individuals in the population. Let $\boldsymbol{Z} = (Z_1, \ldots, Z_n)$ be the vector of treatment assignments in the population of size n, where $Z = 1$ denotes vaccine and $Z = 0$ denotes control. Then the potential outcome of individual i if the population receives treatment assignment $\boldsymbol{Z}$ is denoted by $Y_i(\boldsymbol{Z})$. Halloran and Struchiner (1995) defined the individual direct causal effect of being vaccinated compared with not being

vaccinated in an individual i when the rest of the population $j \neq i$ receives treatment assignment $\boldsymbol{Z}_{j \neq i}$ as

$$Y_i(\boldsymbol{Z}_{j \neq i}, Z_i = 1) - Y_i(\boldsymbol{Z}_{j \neq i}, Z_i = 0). \tag{13.10}$$

The direct causal effect is a family of values that depends on the treatment assignment vector $\boldsymbol{Z}$ in the population.

To define the indirect, total, and overall effects of one vaccination strategy compared with another, one needs to consider a second strategy, denoted $\boldsymbol{Z}'$. They define the individual indirect causal effect of intervention program $\boldsymbol{Z}$ compared with $\boldsymbol{Z}'$ as

$$Y_i(\boldsymbol{Z}_{j \neq i}, Z_i = 0) - Y_i(\boldsymbol{Z}'_{j \neq i}, Z_i = 0), \tag{13.11}$$

where now the individual of interest has not received the vaccine under either intervention programs. Halloran and Struchiner (1995) defined the individual total and overall causal effects analogously. However, they found problems with taking the usual approach in causal inference to average over the potential outcomes to arrive at causal estimands of direct, indirect, total, and overall effects.

Hudgens and Halloran (2008) defined causal estimands of direct, indirect, total, and overall effects in the presence of interference by positing a population of groups, blocks or clusters composed of individuals with interference within the groups but not between the groups as in the study designs described in this chapter. Taking as their point of departure the individual causal effects proposed by Halloran and Struchiner (1995), Hudgens and Halloran (2008) define average individual, group, and population outcomes over all possible treatment assignments for a particular allocation strategy or strategies of interest within and across groups (Sobel 2006). They define causal estimands of the direct, indirect, overall and total effects that are also averages within the groups and across the population of groups. By specifying an assignment mechanism at two levels, that is randomization of groups to allocation strategies, and then randomization of individuals within groups to treatment by the allocation strategy assigned to the group, the average causal direct, indirect, total and overall effects are estimable from the observed outcomes.

The development of the causal estimands is not specific to infectious diseases, and the causal effects are defined based on differences, not relative risks as efficacy measures. For example, consider the data from Ali et al (2005) in Table 13.5. Suppose that the groups with $> 51\%$ and $< 28\%$ coverage are thought of as groups A and B. Effects of vaccination can be estimated based on differences in the incidence of cholera during the first year of follow-up of the trial. The direct effects are estimated by comparing the incidence (risk per 1000 population) between vaccinated individuals and unvaccinated individuals within each group. For example, the estimated direct effect in group B is $7.01 - 2.66 = 4.35$, suggesting vaccination results in 4.35 fewer cases of cholera per 1000 individuals per year. The estimated direct effect in group

A is $1.47 - 1.27 = 0.20$, considerably lower than in group B. The estimated indirect effect in the unvaccinated (B−A) is $7.01 - 1.47 = 5.54$. The estimated total effect (B−A) is $7.01 - 1.27 = 5.74$. Note the total effect (B−A) estimate equals the direct effect estimate in group A plus the indirect effect estimate in the unvaccinated (B−A). The overall effect can be estimated by the difference in incidence between the two groups, i.e., $35/8479 - 25/18,623 = 2.79/1000$.

The approach proposed by Hudgens and Halloran (2008) is a group-randomized study. A one-to-one mapping of the causal estimands of direct, indirect, total, and overall effects of Hudgens and Halloran (2008) to the group randomized studies presented in this chapter is the subject of future research. The next and final section of this chapter contains a brief summary of the formal approach in Hudgens and Halloran (2008).

13.8.2 Formalization

Suppose there are $N > 1$ groups of individuals. For $i = 1, \ldots, N$, let n_i denote the number of individuals in group i and let $\boldsymbol{Z}_i \equiv (Z_{i1}, \ldots, Z_{in_i})$ denote the treatments those n_i individuals receive. Assume Z_{ij} is a dichotomous random variable having values 0 or 1 such that $\boldsymbol{Z}_i$ can take on 2^{n_i} possible values. Let $\boldsymbol{Z}_{i(j)}$ denote the $n_i - 1$ subvector of $\boldsymbol{Z}_i$ with the j^{th} entry deleted. The vector $\boldsymbol{Z}_i$ will be referred to as an intervention or treatment *program*, to distinguish it from the individual treatment Z_{ij}. Let $\boldsymbol{z}_i$ and z_{ij} denote possible values of $\boldsymbol{Z}_i$ and Z_{ij}. Define R^j to be the set of vectors of possible treatment programs of length j, for $j = 1, 2, \ldots$; e.g., $R^2 \equiv \{(0,0), (0,1), (1,0), (1,1)\}$. Possible values $\boldsymbol{z}_i$ of $\boldsymbol{Z}_i$ are elements of R^{n_i}. For positive integer n and $k \in \{0, \ldots, n\}$, define R^n_k to be the subset of R^n wherein exactly k individuals receive treatment 1. E.g., $\sum_{j=1}^{n_i} z_{ij} = k$ for all $\boldsymbol{z}_i \in R^{n_i}_k$.

Denote the potential outcome of individual j in group i under treatment $\boldsymbol{z}_i$ as $Y_{ij}(\boldsymbol{z}_i)$. The notation $Y_{ij}(\boldsymbol{z}_i)$ allows for the possibility that the potential outcome for individual j may depend on another individual's treatment assignment in group i, but the potential outcomes for individuals in group i do not depend on treatment assignments of individuals in group i' for $i' \neq i$.

Treatment Assignment Mechanisms

Let ψ and ϕ denote parameterizations which govern the distribution of $\boldsymbol{Z}_i$ for $i = 1, \ldots, N$. For example, ψ might correspond to randomly assigning half of individuals in a group to treatment 1 and the other half to treatment 0, while ϕ might correspond to assigning all individuals in a group to treatment 0. The goal is to assess the causal effects of assigning groups to to the individual treatment assignment strategy ψ compared to ϕ.

The experimental design is a two-stage randomization procedure. In the first stage, each of the N groups is randomly assigned to either ϕ or ψ. In the second stage, individuals are randomly assigned treatment conditional on their group's assignment in the first stage. For example, in the first stage half

of the N groups might be assigned to an allocation strategy ϕ and the other half ψ; in the second stage, 2/3 of the individuals within a group are randomly assigned treatment 1 for groups assigned ϕ, while 1/3 of the individuals within a group are randomly assigned treatment 1 for groups assigned ψ.

Corresponding to the first stage of randomization, let $\boldsymbol{S} \equiv (S_1, \ldots, S_N)$ denote the group assignments with $S_i = 1$ if the i^{th} group is assigned to ψ and 0 otherwise. Let ν denote the parameterization that governs the distribution of $\boldsymbol{S}$ and let $C \equiv \sum_i S_i$ denote the number of groups assigned ψ.

Average potential outcomes

Similar to Halloran and Struchiner (1995), Hudgens and Halloran (2008) begin by writing the potential outcomes for individual j in group i under $z_{ij} = z$ as

$$Y_{ij}(\boldsymbol{z}_{i(j)}, z_{ij} = z), \tag{13.12}$$

for $z = 0, 1$. They then proceed to define the *individual average potential outcome* under treatment assignment z by

$$\overline{Y}_{ij}(z;\psi) \equiv \sum_{\boldsymbol{\omega} \in R^{n_i - 1}} Y_{ij}(\boldsymbol{z}_{i(j)} = \boldsymbol{\omega}, z_{ij} = z)\Pr\ (\boldsymbol{Z}_{i(j)} = \boldsymbol{\omega} | Z_{ij} = z).$$

In other words, the individual average potential outcome is the conditional expectation of $Y_{ij}(\boldsymbol{Z}_i)$ given $Z_{ij} = z$ under assignment strategy ψ. Averaging over individuals, they define the *group average potential outcome* under treatment assignment z as $\overline{Y}_i(z;\psi) \equiv \sum_{j=1}^{n_i} \overline{Y}_{ij}(z;\psi)/n_i$. Finally, averaging over groups, they define the *population average potential outcome* under treatment assignment z as $\overline{Y}(z;\psi) \equiv \sum_{i=1}^{N} \overline{Y}_i(z;\psi)/N$.

They define the *marginal* individual average potential outcome by $\overline{Y}_{ij}(\psi) \equiv \sum_{\boldsymbol{z} \in R^{n_i}} Y_{ij}(\boldsymbol{z})\ \Pr\ (\boldsymbol{Z}_i = \boldsymbol{z})$, i.e., the average potential outcome for individual j in group i when group i is assigned ψ. Similarly, they define the marginal group and population average potential outcomes by $\overline{Y}_i(\psi) \equiv \sum_{j=1}^{n_i} \overline{Y}_{ij}(\psi)/n_i$ and $\overline{Y}(\psi) \equiv \sum_{i=1}^{N} \overline{Y}_i(\psi)/N$.

Causal estimands

Formally, following Halloran and Struchiner (1995) as in expression (13.11), Hudgens and Halloran (2008) define the *individual direct causal effect* of treatment 0 compared to treatment 1 for individual j in group i by

$$CE^{D}_{ij}(\boldsymbol{z}_{i(j)}) \equiv Y_{ij}(\boldsymbol{z}_{i(j)}, z_{ij} = 0) - Y_{ij}(\boldsymbol{z}_{i(j)}, z_{ij} = 1). \tag{13.13}$$

The causal estimands are then defined in terms of these various average potential outcomes. Hudgens and Halloran (2008) next define the *individual average direct causal effect* for individual j in group i by by

$$\overline{CE}_{ij}^{D}(\psi) \equiv \overline{Y}_{ij}(0;\psi) - \overline{Y}_{ij}(1;\psi), \tag{13.14}$$

i.e., the difference in individual average potential outcomes when $z_{ij} = 0$ and when $z_{ij} = 1$ under ψ. Finally, define the *group average direct causal effect* by $\overline{CE}_{i}^{D}(\psi) \equiv \overline{Y}_{i}(0;\psi) - \overline{Y}_{i}(1;\psi) = \sum_{j=1}^{n_i} \overline{CE}_{ij}^{D}(\psi)/n_i$ and the *population average direct causal effect* by $\overline{CE}^{D}(\psi) \equiv \overline{Y}(0;\psi) - \overline{Y}(1;\psi) = \sum_{i=1}^{N} \overline{CE}_{i}^{D}(\psi)/N$.

Similar to expression (13.11), Hudgens and Halloran (2008) define the *individual indirect causal effect* of treatment program $\boldsymbol{z}_i$ compared with $\boldsymbol{z}_i'$ on individual j in group i by

$$CE_{ij}^{I}(\boldsymbol{z}_{i(j)}, \boldsymbol{z}_{i(j)}') \equiv Y_i(\boldsymbol{z}_{i(j)}, z_{ij} = 0) - Y_i(\boldsymbol{z}_{i(j)}', z_{ij}' = 0), \tag{13.15}$$

where $\boldsymbol{z}_i'$ is another n_i dimensional vector of individual treatment assignments. (Note $\boldsymbol{z}_i'$ does not denote the transpose of $\boldsymbol{z}_i$).

Similar to direct effects, they define the *individual average indirect causal effect* by $\overline{CE}_{ij}^{I}(\phi,\psi) \equiv \overline{Y}_{ij}(0;\phi) - \overline{Y}_{ij}(0;\psi)$. Clearly if $\psi = \phi$, then $\overline{CE}_{ij}^{I}(\phi,\psi) = 0$, i.e., there are no individual average indirect causal effects. Finally, they define the *group average indirect causal effect* as $\overline{CE}_{i}^{I}(\phi,\psi) \equiv \overline{Y}_{i}(0;\phi) - \overline{Y}_{i}(0;\psi) = \sum_{j=1}^{n_i} \overline{CE}_{ij}^{I}(\phi,\psi)/n_i$ and the *population average indirect causal effect* as $\overline{CE}^{I}(\phi,\psi) \equiv \overline{Y}(0;\phi) - \overline{Y}(0;\psi) = \sum_{i=1}^{N} \overline{CE}_{i}^{I}(\phi,\psi)/N$.

Define the *individual total causal effects* for individual j in group i as

$$CE_{ij}^{T}(\boldsymbol{z}_{i(j)}, \boldsymbol{z}_{i(j)}') \equiv Y_{ij}(\boldsymbol{z}_{i(j)}, z_{ij} = 0) - Y_{ij}(\boldsymbol{z}_{i(j)}', z_{ij}' = 1). \tag{13.16}$$

Then define the individual average, group average, and population average total causal effect similar to the indirect causal estimands.

Hudgens and Halloran (2008) define the *individual overall causal effect* of treatment $\boldsymbol{z}_i$ compared to treatment $\boldsymbol{z}_i'$ for individual j in group i by $CE_{ij}^{O}(\boldsymbol{z}_i, \boldsymbol{z}_i') \equiv Y_{ij}(\boldsymbol{z}_i) - Y_{ij}(\boldsymbol{z}_i')$. Similarly, for the comparison of ϕ to ψ, define the *individual average overall causal effect* by $\overline{CE}_{ij}^{O}(\phi,\psi) \equiv \overline{Y}_{ij}(\phi) - \overline{Y}_{ij}(\psi)$, the *group average overall causal effect* by $\overline{CE}_{i}^{O}(\phi,\psi) \equiv \overline{Y}_{i}(\phi) - \overline{Y}_{i}(\psi)$ and the *population average overall causal effect* by $\overline{CE}^{O}(\phi,\psi) \equiv \overline{Y}(\phi) - \overline{Y}(\psi)$.

Estimation and inference

Assuming the randomized assignment strategies at both levels of randomization in which the number of groups randomized to a strategy is fixed, and the number of individuals within each group randomized to received treatment is fixed, Hudgens and Halloran (2008) show that the observed data yield unbiased estimators of the causal estimands. Suppose $S_i = 1$. They show that

$$\widehat{Y}_i(z;\psi) \equiv \frac{\sum_{j=1}^{n_i} Y_{ij}(\boldsymbol{Z}_i) I[Z_{ij} = z]}{\sum_{j=1}^{n_i} I[Z_{ij} = z]} \text{ for } z = 0, 1, \tag{13.17}$$

i.e., $\widehat{Y}_i(z;\psi)$ is the average of observed outcomes for individuals in group i receiving treatment z under treatment program $\boldsymbol{Z}_i$ is an unbiased estimator of $\overline{Y}_i(z;\psi)$ Also, $\widehat{\overline{CE}_i^{D}}(\psi) \equiv \widehat{Y}_i(0;\psi) - \widehat{Y}_i(1;\psi)$ is a conditionally unbiased estimator of $\overline{CE}_i^{D}(\psi)$ given $S_i = 1$. Finally, they show that or $z = 0, 1$, let $\widehat{Y}(z;\psi) \equiv \sum_{i=1}^{N} \widehat{Y}_i(z;\psi) I[S_i = 1] / \sum_{i=1}^{N} I[S_i = 1]$. is an unbiased estimator of $\overline{Y}(z;\psi)$ for $z = 0, 1$. Thus, unbiased estimators for the population average direct, indirect, and total causal effects are given by $\widehat{\overline{CE}}^{D}(\psi) \equiv \widehat{Y}(0;\psi) - \widehat{Y}(1;\psi)$, $\widehat{\overline{CE}}^{I}(\phi,\psi) \equiv \widehat{Y}(0;\phi) - \widehat{Y}(0;\psi)$, and $\widehat{\overline{CE}}^{T}(\phi,\psi) \equiv \widehat{Y}(0;\phi) - \widehat{Y}(1;\psi)$ where $\widehat{Y}(z;\phi)$ is defined analogously to $\widehat{Y}(z;\psi)$ for $z = 0, 1$.

Let $\widehat{Y}_i(\psi) \equiv \sum_{j=1}^{n_i} Y_{ij}(\boldsymbol{Z}_i)/n_i$ and $\widehat{Y}(\psi) \equiv \sum_{i=1}^{N} \widehat{Y}_i(\psi) I[S_i = 1] / \sum_{i=1}^{N} I[S_i = 1]$. Hudgens and Halloran (2008) show that the unbiased estimator of the overall effect causal estimand $\overline{CE}^{O}(\phi,\psi)$ is given by $\widehat{\overline{CE}}^{O}(\phi,\psi) \equiv \widehat{Y}(\phi) - \widehat{Y}(\psi)$ where $\widehat{Y}(\phi)$ is defined analogously to $\widehat{Y}(\psi)$.

With a further assumption of stratified interference, that is, that the potential outcomes depend on the number within a group that receives a treatment, but not exactly which ones, they derive variance estimators.

Problems

13.1. Constrained randomization
(a) Consider designing a community-randomized trial of a cholera vaccine in six communities. The average annual incidence of cholera in the six communities is 1, 3, 4, 9, 10, 12%. How many different allocations of the vaccine and control are there for a pair-matched design? For a completely randomized (at the group level) design?
(b) What would be a reasonable constraint to ensure a fairly balanced allocation under complete randomization?

13.2. Overall effectiveness in Alaska of PCV7
(a) serotypes age 2–4 years in Alaska Natives and non-Natives
(b) The second part of the problem is described here.

13.3. Computing sample size in stepped wedge design
(a) Assume $Z_E Z_{SW} = 1.2$ Suppose one wants to have power of 80% and Type I error of 5% in a study. What values of z-score does one need to use in does one need to use in expression 13.3
(b) By what factor would the sample size in expression 13.3 be multiplied? (see Moulton et al 2007, p 195)
(c) Suppose one decided to change the entry of groups into the intervention arm at every three months instead of every two months. What would be the approximate increase in effective sample size barring any substantial secular trends?