Chapter 6
Evaluating Protective Effects of Vaccination

6.1 Overview

Evaluating the direct protective effects of vaccines in the individuals who were vaccinated has been the focus of vaccine studies over the past century. Generally, interest has been in the ability of vaccination to prevent or to ameliorate disease rather than to prevent infection (Clements-Mann 1998). Ascertainment of cases is often done by finding suspected cases in the population under study in people who exhibit a set of symptoms. The suspected cases are then tested for biological confirmation of the infectious agent of interest. Alternatively, surveillance can ascertain cases reported in central registries. However they are ascertained, with most vaccines, clinical disease is the primary outcome of interest. When ascertainment is on clinical cases, most asymptomatic infections may go undetected. A different situation arises when infection is the primary outcome. To ascertain infections in asymptomatic people, an active follow-up method of testing asymptomatic people is needed.

In this chapter we consider estimation and inference for direct protective effects of vaccination, VE_S and VE_{SP}, in studies that do not condition on exposure to infection. We consider aspects of the design of such studies. Several examples of randomized, double-blind (double-masked) controlled vaccine trials illustrate the standard approach to design and analysis of such studies. Our choice of studies to present was motivated largely because of their use as illustrations in other sections of the book. Most randomized and pivotal studies of vaccines have been based on VE_{SP} or VE_S. Much has been written on studies to meet the approval of the regulatory agencies, and the design of clinical trials generally. Our goal here is to consider how VE_S and VE_{SP} relate to other measures of vaccine effects within the dependent happening context, and consider a few design considerations. Because VE_S and VE_{SP} do not condition on exposure to infection, assumptions about the relative exposure opportunity in the vaccine and control groups are important.

We have generally distinguished VE_S, the vaccine efficacy for susceptibility to infection, from VE_{SP}, the vaccine efficacy for susceptibility to disease. However, in this and the following chapters, ascertainment is most often on disease rather

M.E. Halloran et al., *Design and Analysis of Vaccine Studies*, Statistics for Biology and Health, DOI 10.1007/978-0-387-68636-3_6,

than infection. In both instances, the population at risk is individuals susceptible to infection. In a sense, one can imagine a continuum after randomization that includes infection, development of symptoms, and possibly development of severe disease. Most of the methods apply equally well if ascertainment is on infection or clinical disease. In this and the next two chapters, we use VE_S often to denote situations where the primary outcome can be either infection or disease. Which one is meant is clear from the context. Any outcome that is the first cut after randomization will provide a statistically valid assessment of the effect of the vaccine on that outcome.

This is in contrast to VE_P, the vaccine efficacy for progression or post-infection outcomes. In this situation, the vaccine effect of interest is in an outcome that occurs only in those people who become infected. The methods of analysis and potential for biases are different for VE_P. In Chapter 7, we discuss different conceptual models of protective effects of vaccine and the consequences for choosing and interpreting protective efficacy estimates. The chapter also discusses methods to estimate waning vaccine effects. In Chapter 8 we present further topics in evaluating protective effects. The evaluation of the effect of vaccination on post-infection outcomes is considered in Chapter 9.

6.2 Estimating VE_S

The vaccine efficacy measures of interest in this chapter are the Levels II, III, and IV parameters in Table 2.2 that do not condition on exposure to infection. The Level IV measure $\mathrm{VE}_{S,CI}(T)$ is defined using the cumulative incidence or attack rates at the end of the study:

$$\begin{aligned} VE_{S,CI}(T) &= 1 - \frac{\text{vaccinated infection events/persons–at–risk}}{\text{unvaccinated infection events/persons–at–risk}} \\ &= 1 - \frac{CI_1(T)}{CI_0(T)}\,. \end{aligned} \tag{6.1}$$

The Level II parameters $\mathrm{VE}_{S,IR}$ based on the incidence rates and $\mathrm{VE}_{S,\lambda}$ based on the hazard rates require knowledge of the infection times:

$$\begin{aligned} VE_{S,IR}(T) &= 1 - \frac{\text{vaccinated events/person-time}}{\text{unvaccinated events/person-time}} \\ &= 1 - \frac{IR_1(T)}{IR_0(T)}. \end{aligned} \tag{6.2}$$

The $\mathrm{VE}_{S,\lambda}$ based on the hazard rate ratio is

$$\mathrm{VE}_{S,\lambda}(t) = 1 - \frac{\lambda_1(t)}{\lambda_0(t)}. \tag{6.3}$$

The Level III parameter VE_{PH} based on the proportional hazards model requires only the ordering of the infection times:

$$VE_{S,PH} = 1 - \exp(\beta). \tag{6.4}$$

where β is the log hazard ratio. In Chapter 2 we showed the intrinsic relationship of the parameters to one another based on the dependent happening relation (2.7). We also showed they form a hierarchy based on the amount of information required for their estimation.

In this chapter, we treat $VE_{S,IR}(T)$, $VE_{S,\lambda}$, and $VE_{S,PH}$ somewhat interchangeably. The interpretation of $VE_{S,CI}(T)$ and $VE_{S,IR}(T)$ ($VE_{S,\lambda}$, $VE_{S,PH}$) differ substantially. $VE_{CI}(T)$ is related to the number of cases saved over the period of the study, and $VE_{IR}(T)$ and the other two parameters measure a relative improvement in incidence rate or hazard, whereby both are underestimates if dependent happenings are not taken into account (Section 2.8.1). The choice between $VE_{CI}(T)$ and a vaccine efficacy based on incidence or hazard ratios could be influenced by the distribution of vaccine protection (Chapter 7).

6.2.1 Absolute versus relative efficacy

The control arm in a planned study is often another active vaccine assumed not to have an effect on the disease of interest. In the pneumococcal conjugate vaccine study below, a meningococcal conjugate vaccine is the control. In these studies, the goal is to show that the active vaccine of interest is superior to the control in preventing the primary outcome of interest. If a licensed (and recommended) vaccine is available for the disease of interest, it is generally unethical to use a placebo or vaccine against a different disease in the control arm. Then the study must compare two (or more) active vaccines against the same disease. The relative rather than the absolute efficacy can be computed. The relative efficacy is the relative reduction in disease risk or incidence by the one vaccine compared with the other. An example is the pertussis vaccine study in Senegal presented below. The whole cell pertussis vaccine was recommended for infants in Senegal, so the acellular pertussis vaccine could not be compared to a placebo. In contrast, in Sweden, the whole cell pertussis vaccine had been discontinued, so there was no licensed pertussis vaccine in Sweden when they conducted the study of the acellular pertussis vaccine. In the Swedish study, the control was the diphtheria–tetanus toxoid without the pertussis component.

As new generations of vaccines are introduced, it is more common to be comparing a new vaccine candidate with an existing vaccine. If both vaccines are fairly efficacious and or the outcome of interest is fairly rare, then the size of the field study becomes prohibitively large and expensive. For example, the pneumococcal vaccines are highly efficacious against invasive disease, so that field studies of new pneumococcal vaccines with invasive disease as the primary outcome are not fea-

sible. In this setting, the hunt for immunological surrogates of protection becomes imperative. In the case of pneumococcal vaccines, there is also interest in developing pneumococcal nasopharyngeal carriage as a primary outcome for vaccine field study (Chapter 15).

Even when individuals cannot be randomized to a placebo, there may be individuals under surveillance who do not enroll in the trial, and thus do not receive either vaccine. The absolute efficacy of both vaccines can be computed by comparison with the individuals who happened not to be in either study arm. The study is then an observational cohort study, not a randomized study. The Senegal pertussis vaccine study included surveillance of cases in people not in the study, so was able to compute the absolute efficacy of both vaccines, although with the potential biases inherent in observational studies. With two active vaccines, the trial may be planned in a way to show that the efficacy of the new vaccine is not worse than the already licensed vaccine (a noninferiority study) or that the new vaccine has a higher efficacy than the other vaccine (a superiority trial), the usual approach in vaccine trials that compare a vaccine to a control.

6.2.2 Types of studies

Cohort studies for evaluating vaccines follow groups of people over time, some of whom are vaccinated, some of whom are not. Randomized vaccine studies are examples of cohort studies in which the vaccine has been randomly allocated. Cohort studies can be used to estimate any of the unconditional VE_S parameters if certain conditions are met. If all of the vaccine was administered before the beginning of the observation period, then the cohort is a fixed cohort. If, in addition, there is no loss to follow-up during the observation period, the cohort is a closed cohort. Then $\mathrm{VE}_{S,CI}(T)$ can be estimated by the cumulative incidence or attack rates. More generally, open or dynamic cohorts allow people to join and leave the population under study and to change their vaccination status. From these studies in dynamic cohorts, estimates can be based on either cases per person-time at risk, the incidence rate, or using survival analysis methods in which the risk set can change over time. $\mathrm{VE}_{S,IR}$ and $\mathrm{VE}_{S,\lambda}$ can be estimated from either closed or open cohorts. Primary vaccine efficacy studies often report $\mathrm{VE}_{S,IR}$ based on relative events per person-time, or Level II information.

In a case-control study, cases are ascertained and controls selected from a source population. The goal of the case-control study is to estimate the same unconditional estimands of vaccine efficacy as in the cohort studies. The method of sampling the controls and the method of analysis determine whether the case-control study will provide good estimates for $\mathrm{VE}_{S,IR}$, $\mathrm{VE}_{S,\lambda}$, or $\mathrm{VE}_{S,CI}(T)$. A case-control study can be thought of as a sample of data from a hypothetical cohort study. The cohort can also be thought of as a source population that gives rise to the cases (Chapter 8).

6.2.2.1 Randomized versus observational cohort studies

Greenwood and Yule (1915) stated three conditions for valid inference in vaccine studies:

1. The persons must be, *in all material respects,* alike.
2. The effective exposure to the disease must be identical in the case of inoculated and uninoculated persons.
3. The criteria of the fact of inoculation and of the fact of the disease having occurred must be independent.

The conditions for a valid comparison are essentially met under randomization. Randomization is supposed to ensure that potential confounders are balanced between the two groups. Observational studies that do not assign vaccine randomly need to examine the three criteria carefully. The criteria can be thought of in terms of exposure to infection versus susceptibility to infection. First, randomization is supposed to ensure that the groups being compared are in all relevant aspects alike. Relevant covariates can include pre-existing immune levels such as antibody titer, prior vaccination, prior disease history, age, and gender, among others.

Second, randomization is supposed to ensure that effective exposure to infection of the two groups is the same. The two groups having the same exposure to infection is not the same as every person in the groups having the same exposure to infection. Even if, on average, exposure in the two groups is comparable, there may be heterogeneity of exposure to infection within the groups. Some participants might not be exposed at all to the infectious agent of interest. Because in field trials, exposure to infection is not under control of the investigator, in studies that do not condition on exposure to infection, the assumption of equal exposure in the two groups is a strong one, especially if a study is not randomized. For example, children of a higher socioeconomic status may be less exposed to a certain infection. If these children also tend to get vaccinated, then a study of the effect of vaccination will overestimate vaccine efficacy. Potential relevant covariates related to exposure to infection could include distance from potential environmental sources of infection, number of people living in the household, use of bednets, behavioral covariates such as number of sexual contacts or handwashing habits, among others. Going to work rather than working at home or attending school rather than either being too young to attend school or remaining at home for other reasons can affect exposure to infection.

Third, the chance of being vaccinated cannot be associated with the probability of developing disease. Some of these elements are similar to those in the first group related to susceptibility to infection and disease. As an example, children of a higher socioeconomic status may have better nutrition, and therefore better immune systems and better resistance to infection or disease if exposed. If children of higher socioeconomic status also tend to be vaccinated, then a study of the effect of vaccination will overestimate the vaccine efficacy. In both of these situations, socioeconomic status could be used as a proxy covariate for either exposure or for susceptibility to infection.

Table 6.1 Number of individuals, number of cases, and number of person-time at risk in vaccinated and control groups

	Number of Persons in Group	Number of Cases of Disease	Person-Time at Risk
Vaccinated	N_1	c_1	Y_1
Control	N_0	c_0	Y_0

If these three criteria are met, any differences in the rate of developing disease in the two groups is likely due to the biological effects of the vaccine. It is important to collect information on relevant covariates and potential confounders in both randomized and observational studies. Potential confounders will depend on the particular infectious agent of interest and the setting of the study. Reports of randomized cohort and case-control studies usually include a comparison of the vaccine and control groups on any covariates considered relevant.

Nonrandomized cohort and case-control studies need to address these potential sources of bias. Although propensity scores (Rosenbaum 1995) and marginal structural models (Robins et al 2000a) could be used to adjust for confounding in vaccine studies, these approaches have not found much use thus far. Further details of epidemiologic study design can be found in Rothman et al (2008). Interactions of preexisting immunity and level of exposure to infection can confound interpretation of vaccine efficacy estimates even when the study is randomized (Chapter 14).

6.2.3 Estimation and inference

The statistical methods for analyzing the studies described in this chapter are fairly standard. Consistent with the philosophy of this book, estimation with a measure of uncertainty such as confidence intervals, likelihood intervals, or a Bayesian posterior distribution is the focus rather than hypothesis testing. Our interest is in the estimate of vaccine efficacy and the interpretation of the estimate. Consider a vaccine study with N_1 individuals in the vaccine group and N_0 in the control group, and $N = N_0 + N_1$. The cohort can be observed either at time 0 and time T or over the interval $[0, T]$. The number of cases observed in the unvaccinated group is c_0 and in the vaccinated group is c_1. The total person-time at risk in each group is denoted by Y_0 in the unvaccinated group and Y_1 in the vaccinated group (Table 6.1).

Estimating $\mathrm{VE}_{S,CI}(T)$ based on the cumulative incidence or attack rates requires only information about whether persons are infected by the end of the study at time T, that is, final value data:

$$\mathrm{VE}_{S,CI}(T) = 1 - \frac{c_1/N_1}{c_0/N_0} \,. \tag{6.5}$$

Estimation of $VE_{S,CI}(T)$ based on the simple relative proportions of cases in each group assumes that there is no loss to follow-up, that is, no censoring.

Chick et al (2001) consider correcting for bias in risk ratio and vaccine effect estimators, especially when the number of cases is small. The standard maximum likelihood vaccine effect estimators are consistent, but they are biased because they are nonlinear functions of other estimators. The bias is small when the number of cases is relatively large, say >70 in the placebo arm. However, with small numbers of cases, the bias can be substantial. Chick et al (2001) propose various bias correction options, including one suggested by Jewell (1986). Bias of both the $VE_{S,CI}$ under an all-or-none model and the $VE_{S,CI}$ under the leaky model are explored. Of the options considered, the best was to add one to the positive count in the control population, both to the case count and the population count. For example, they recommend using

$$\widehat{VE}_{S,CI}(T) = 1 - \frac{c_1/N_1}{(c_0+1)/(N_0+1)}. \tag{6.6}$$

This addition in the control population increases the $VE_{S,CI}(T)$ estimates. As they point out, it may seem to "corrupt the data." However, for small studies, the simulations are convincing. Clearly when c_0 is large, the addition of one count will have a small effect. They also provide bias corrections for Bayesian vaccine effect estimators, for VE_I and VE_S based on the secondary attack rates, and for the vaccine effect of the susceptibility and infectiousness effects on the reproductive number.

$VE_{S,IR}(T)$ based on relative incidence rates is estimated by

$$VE_{S,IR}(T) = 1 - \frac{c_1/Y_1}{c_0/Y_0}. \tag{6.7}$$

The usual assumption is that the numbers of events follow a Poisson distribution. Similarly, from time-to-event data, to estimate $VE_{S,\lambda}$ investigators may estimate the instantaneous hazard rates in the vaccinated and unvaccinated $\lambda_1(t)$ and $\lambda_0(t)$, respectively, using survival analysis methods. When covariates such as age and gender are added, the analyses are stratified by the covariates or Poisson regression can be used.

Under the assumption that the effect of the vaccine is multiplicative, constant, and homogeneous, the Cox proportional hazards model can be used to estimate $VE_{S,PH}$. In this case, it is not necessary to estimate the hazard rate in the unvaccinated group, but only the relative hazard rate. Covariates including time-dependent covariates can easily be incorporated using standard software. The proportional hazards model with covariates can be used to investigate possible confounding factors. Because the proportional hazards model assumes that the baseline hazard is the same in both the vaccinated and the unvaccinated groups, for studies including different communities, it may be possible to include a covariate for each community. The model could then assume that the incidence varies by community, but the vaccine effect is the same in each community (Section 6.4.1).

Several approaches are available for the confidence interval for $\mathrm{VE}_{S,CI}(T)$. O'Neill (1988) favored the method based on the log of the ratio of two binomial random variables (Katz et al 1978) because of its simplicity of interpretation and the symmetry of the confidence interval on the log scale. Let $\theta(T) = CI_1(T)/CI_0(T)$, so that $\mathrm{VE}_{S,CI}(T) = 1 - \theta(T)$, and let $\beta(T) = \ln\theta(T)$. Assume for now that the follow-up is over the interval T, so that we can drop the T from the notation. The estimate of θ is $\hat{\theta} = (c_1/N_1)/(c_0/N_0)$ and $\hat{\beta} = \ln\hat{\theta}$. An estimate of the variance of β is

$$\sigma^2 = \frac{N_1 - c_1}{N_1 c_1} + \frac{N_0 - c_0}{N_0 c_0} = \frac{1}{c_1} + \frac{1}{N_1} + \frac{1}{c_0} + \frac{1}{N_0}. \tag{6.8}$$

In vaccine studies, N_0 and N_1 are usually large, so that the variance of β is approximated by a function of the number of cases in the vaccinated and unvaccinated groups, $1/c_1 + 1/c_0$. The $100(1-\alpha)$ percent confidence interval for $\mathrm{VE}_{S,CI}(T) = 1 - \theta$ is

$$[1 - \exp(\hat{\beta} + z\hat{\sigma}), 1 - \exp(\hat{\beta} - z\hat{\sigma}], \tag{6.9}$$

where z is the $(1-\alpha)$ percentage point of the standard normal distribution. One can also use Taylor series approximations (Hightower 1988). Ewell (1996) compared Bayesian posterior regions with frequentist exact and large sample confidence intervals for intermediate (Phase IIb) trials. Koopman's (1984) method for the ratio of two binomials is also used. Generally two-sided intervals are recommended, and even required by some journals. The lower confidence bound on the vaccine efficacy estimate is sometimes of primary interest, especially in proof-of-concept studies, or Phase IIb studies.

An approximate confidence interval for $\mathrm{VE}_{S,IR}$ can be obtained similarly as in (6.9). An estimate of the approximate variance of the log of the ratio of the incidence rate in the vaccinated group and the incidence in the unvaccinated group is again

$$\hat{\sigma}^2 = \frac{1}{c_1} + \frac{1}{c_0}. \tag{6.10}$$

If now $\theta = (c_1/Y_1)/(c_0/Y_0)$ and $\beta = \ln\theta$, then the $100(1-\alpha)$ percent confidence interval for $\mathrm{VE}_{IR} = 1 - \theta$ is

$$[1 - \exp(\hat{\beta} + z\hat{\sigma}), 1 - \exp(\hat{\beta} - z\hat{\sigma}], \tag{6.11}$$

where z is the $(1-\alpha)$ percentage point of the standard normal distribution.

If there is loss to follow-up, then $\mathrm{VE}_{S,CI}(T)$ also requires knowledge of the time of onset of cases. In a hepatitis B vaccine study, Szmuness et al (1980) calculated cumulative attack rates using a life-table method. The statistical significance of the differences between observed numbers of trial endpoints in different groups was calculated from the life tables by the log-rank summary chi-square test. In another hepatitis B vaccine study, Francis et al (1982) also used a life-table approach based on person-months of follow-up to get cumulative attack rates. Hudgens et al (2004)

suggest using nonparametric maximum likelihood estimators of CI_1 and CI_0 in the presence of censoring (Kaplan and Meier 1958; Peto 1973). Standard survival analysis methods can be used for inference for $VE_{S,\lambda}$ and $VE_{S,PH}$:

$$\widehat{VE}_{S,PH} = 1 - \exp(\hat{\beta}). \tag{6.12}$$

where $\hat{\beta}$ is the partial likelihood estimate of the log hazard ratio (Cox 1972). The methods for the above analyses in this chapter are available on most statistical analysis packages.

When the number of cases in the study is small, exact confidence intervals may be used. Again, many approaches are available for exact confidence intervals. Randomized trials in this chapter used the Clopper–Pearson (1934) or Koopman's (1984) method. Agresti and Coull (1998) compare exact and approximate confidence intervals and find that sometimes approximate intervals are better than exact. Specialized software is available for most exact computations.

6.3 Design Considerations

In this section, we consider some of the design considerations of a vaccine study, with the studies in the next section serving as illustrations.

6.3.1 Vaccines and vaccination schedule

The vaccine of interest and the comparison, whether active control, placebo, or nothing, need to be specified. If active administration of vaccines is part of the study design, the number of doses, and the schedule for administering the doses need to be specified. Many vaccines require two or more doses for complete vaccination. For example, usually complete pertussis vaccination requires three doses. It is important when possible to record the number of doses of a vaccine that a person has received to determine if the person has complete or incomplete vaccination. In addition, the immune response requires some time to develop. Thus, many studies include only cases in the analysis that occur a certain time interval after the completion of vaccination. In a randomized study, participants who receive the number of doses according to protocol are included in the per-protocol analysis. In the intent-to-treat analysis, any person randomized to a particular arm regardless of how many doses received is included in the analysis. Analyses can also be broken down by the actual number of doses received.

In observational studies, the study can specify what the recommended dose schedule is for that vaccine, then ascertain the extent to which participants are vaccinated according to the recommended schedule.

6.3.2 Study population

The study needs to specify the usual person, time, and place of any field study, whether randomized or observational. Eligibility and exclusion criteria need to be specified.

6.3.2.1 Recruitment and vaccination

Recruitment into a vaccine study can be through a population-based study, a local census, by attendance at clinics or physician's offices, schools, workplaces, health maintenance organizations or public advertisements. The method of recruitment will depend on the societal context and the target age of vaccination. Vaccination can take place in clinics or by teams going to the field for vaccination.

6.3.3 Case definition

The case definition is an essential element for the study. In randomized studies, there will usually be a primary endpoint for the primary analysis. The case definition can be defined by clinical criteria alone or require biological confirmation of evidence of the infectious agent of interest. Several secondary endpoints may be based on different case definitions, other clinical endpoints related to the infectious agent of interest, or laboratory endpoints related to either the immune response or the course of the infection. Hudgens et al (2004) reviewed endpoints in vaccine trials.

6.3.4 Ascertainment of cases

Methods for ascertaining potential clinical cases include active surveillance such as through phone calls at specified intervals or visits to the homes. Suspect cases may be ascertained in clinical settings, whereby only cases that seek medical attention will be ascertained. If the case definition includes biological confirmation, then the relevant tests will be performed. Ascertainment of infected people rather than clinical cases requires testing of all of the study participants at regular intervals.

6.3.4.1 Safety and Immunogenicity

If a study actively administers vaccine, usually study participants will be directly observed for a period of time for short-term adverse events such as anaphylactic reaction. Parents or adults can be given diaries to keep track of adverse events. Investigators may make visits or phone calls to the homes of participants to register

any adverse events. Immunogenicity of the vaccine could be measured on all or a subset of participants. It may not be measured on anyone. In observational studies, immunogenicity measures may not be available.

6.3.5 Sample size calculations

It happens often that vaccine studies go to the field, then suddenly there is no or little transmission, so there are few events. Someone once said that for vaccine studies, one should calculate the sample size then multiply by 5 or possibly 10. Here are a few formulae for simple sample size calculations as guidelines, but most sample size calculations for vaccine studies will need computer simulations. Careful, sometimes lengthy, baseline studies to understand the local epidemiology and transmission of the infection, seasonal and yearly variation in incidence, and other characteristics may be required before sample size calculations can be considered reliable.

Hayes and Bennett (1999) provide simple formulae for individually randomized studies which we summarize here as well as parallel design cluster randomized studies (Chapter 13). Let $z_{\alpha/2}$ 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 vaccine 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 amount of person-time required in each group is (Smith and Morrow 1996; Hayes and Bennett 1999)

$$y = (z_{\alpha/2} + z_\beta)^2 \frac{\lambda_0 + \lambda_1}{(\lambda_0 - \lambda_1)^2}. \tag{6.13}$$

If the outcome is based on proportions, let π_0 and π_1 be the true population proportions in the presence and absence of the intervention. Let n be the number of individuals in each group. Then the number of individuals required in each arm is

$$n = (z_{\alpha/2} + z_\beta)^2 \frac{\pi_0(1-\pi_0) + \pi_1(1-\pi_1)}{(\pi_0 - \pi_1)^2}. \tag{6.14}$$

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

$$n = (z_{\alpha/2} + z_\beta)^2 \frac{\sigma_0^2 + \sigma_1^2}{(\mu_0 - \mu_1)^2}. \tag{6.15}$$

Fay et al (2007) consider sample size calculations for testing differences in means between two samples and allowing for different variances in the two groups. The approach accounts for two sources of variability. One source of variability is in parameter estimates that are estimated from prior data. The second source of variability is if the vaccine fails in some of the people who are vaccinated. The sample size calculation needs to take the possible failure of the vaccine into account. The research was motivated by the design of a Phase II trial of a *Plasmodium falciparum* blood-stage malaria vaccine candidate in Africa. Baseline data on malaria in children had been gathered in a village in Mali in 1999 and 2000. Children were visited weekly and blood smears were done monthly. Data on malaria symptoms and blood smears were available. Several different primary endpoints for the trial were explored. The goal of vaccination was to elicit an immune response comparable to the immune response in older children, all of whom had had repeated exposure to malaria infection. For each candidate primary endpoint, the effect measure was defined as the difference in the malaria outcome in the older compared to the younger children. Instead of choosing an effect size arbitrarily, the observational data were used to estimate the standardized effect size and variances. The variability in the variance estimate can be accounted for simply by using a slightly larger nominal power in the usual sample size calculation, called calibrated power. Fay et al (2007) provide a table of calibrated power by sample size.

The second problem in designing the trial was that some of the children might not respond to the vaccine, for genetic or other reasons, An example would be an all-or-none distribution of protection. For the second problem, the proportion expected not to respond to the vaccine could be obtained from expert opinion, as in traditional sample size computations. Fay et al (2007) provide simple closed form sample size calculations. In general, the sample size will be greater if a proportion of the population does not respond to the vaccine than if all respond to the vaccine.

6.4 Examples of Randomized Trials

6.4.1 Relative efficacy of pertussis vaccines in Senegal

A randomized, double-blind trial comparing a diphtheria–tetanus–acellular pertussis vaccine (DTaP) (pertussis toxoid and filamentous hemagglutinin) with a whole cell vaccine (DTwP) was conducted in the Niakhar area of Senegal (Simondon et al 1997). (See Section 10.2.3 for more details about the area.) The comprehensive ongoing surveillance in the Niakhar area allowed a prospective, nested case-contact study and a cohort study to be conducted during the trial to estimate absolute efficacy of each vaccine.

Eligible infants were those born between February 1, 1990, and April 30, 1994 to mothers residing in the Niakhar area who attended the vaccination sessions. From 1990 through 1994, 4181 children were randomized to receive one of the vaccines

at 2, 4, and 6 months. Surveillance by weekly home visits looked for cough illness persisting more than 7 days in all children under 15 years of age, including children not in the study. Adverse events were screened in the first two weekly visits following each vaccine dose using a standardized questionnaire. Any positive answer was followed up by a physician. The physicians doing the examinations took samples for culture and serological testing blinded to vaccination status. The primary protocol definition of a case of pertussis was defined as 21 or more days of cough confirmed by (a) positive bacterial culture from nasopharyngeal aspirates, (b) serology (IgG against pertussis toxoid and filamentous hemagglutinin), or (c) contact with a culture-confirmed person in the same compound and coughing had started within 28 days before or after onset of illness in the culture-confirmed child (epilink). Polymerase chain reaction (PCR) amplification was used to detect *B. pertussis* DNA in nasopharyngeal aspirates.

The study sample size had been determined assuming that the efficacy of the whole cell vaccine was 75% and allowed detection of the relative ratio of 1.5 in the two arms of the study at the 0.05 significance level. The overall ratio of pertussis incidence in the DTaP group relative to the DTwP group ($RR_{ac/wc}$) and confidence interval were estimated in a proportional hazards model with calendar time as the time scale and stratified by village. Pertussis is epidemic and the proportional hazards model assumes that the baseline hazard is equal in the comparison groups. The model allows the incidence to vary by village, but assumes that the rate ratio is the same across villages. A multivariate proportional hazards model was used to investigate confounding factors. A secondary intent-to-treat analysis included all children receiving at least one dose of the study vaccines. After the study began, the WHO recommended that the case definition be 21 or more days of paroxysmal cough, not just cough. For each child, surveillance ended either at the onset of pertussis, additional pertussis immunization, emigration, death, or refusal to continue in the investigation. All surveillance for the study ended December 31, 1994.

Comparability between children receiving three doses was checked for age at inclusion, gender, weight at first dose, rank of birth number, age of mother, number of persons in the compound, and the number of persons <15 years of age in the compound. No significant differences were found. During the period of surveillance, physicians confirmed at least one episode of >7 days cough in 837 of 2567 compounds reporting such episodes to field workers. The total duration of follow-up was 3165 person-year at risk in the DTwP group and 3193 person-year at risk in the DTaP group. Table 6.2 contains the number of cases and incidence rate ratios for different case definitions. The primary analysis considered cases that occurred $\geq$28 days after the third dose. The overall ratio of pertussis incidence in the DTaP group relative to the DTwP group ($RR_{ac/wc}$) using the protocol case definition was 1.54 (95% CI, 1.23–1.93). A multivariate proportional hazards analysis including the comparability factors revealed that children in compounds with more than 30 members had a higher rate of pertussis, but the value of $RR_{ac/wc}$ did not change.

In a cohort analysis of 229 unvaccinated children, using the same proportional hazards model and the protocol case definition, absolute efficacy was 66% (95% CI, 46–78) for DTwP and 48% (95% CI, 18–66) for DTaP. Using the WHO case

Table 6.2 Incidence rate ratio of DTaP (acellular pertussis) vaccine compared with DTwP (whole cell pertussis) vaccine for different case definitions in the Niakhar, Senegal study (Simondon et al 1997)

	No. of Cases		
	Whole Cell Vaccine	Acellular Vaccine	Incidence Rate Ratio [95% CI]
≥21 days of cough (protocol definition)			
Protocol confirmation criteria	123	197	1.54 [1.23–1.94]
Intention-to-treat	162	233	1.43 [1.16–1.74]
With PCR	65	128	1.87 [1.38–2.52]
≥21 days of paroxysmal cough (WHO definition)			
Protocol confirmation criteria	16	41	2.42 [1.35–4.34]
Intention-to-treat	23	49	2.06 [1.25–3.39]
With PCR	10	31	2.80 [1.36–5.74]

definition, the absolute efficacies were 91% (95% CI, 81–96) for DTwP and 79% (95% CI, 58–89) for DTaP.

This study illustrates several points. First, vaccine studies sometimes report the relative risk or rate ratios rather than the vaccine efficacies. Vaccine efficacy has the awkward property that it ranges from 1 to $-\infty$. The relative risk or rate ratios range from 0 to ∞ with the value of 1 being associated with no relative effect. Second, different case definitions can substantially alter the estimates. In the comparison of the DTaP to DTwP, the point estimates of the rate ratios were higher with the WHO definition, although the confidence intervals overlap. The absolute efficacy of both vaccines in the cohort analysis was higher with the WHO definition. The choice of case definition in pertussis is the subject of ongoing international discussion. The pertussis study in the next section uses a slightly different definition.

6.4.2 Absolute efficacy of pertussis vaccine in Sweden

Because of its limited efficacy, the Swedish-made whole cell pertussis vaccine was withdrawn in 1979. After that, Sweden had no licensed pertussis vaccine, so it was possible to conduct a randomized, placebo-controlled trial (Trollfors et al 1995). Infants were randomly assigned to receive DT toxoids or the same DT toxoids with pertussis toxoid (DTaP toxoids). The vaccine contained only the single component of the pertussis toxoid. About 99% of children in Sweden visit publicly financed child health clinics, where information about the study was given to the parents of infants. Full-term healthy infants in the Göteberg area were eligible if the family had a telephone and at least one parent spoke Swedish. The vaccinations and follow-up were performed at five study sites. The parents of 3450 of 5964 eligible children

Table 6.3 Pertussis vaccine efficacy, $VE_{S,IR}$, of DTaP compared with DT for different case definitions during the main period of follow-up (30 days after the third vaccination until the end of the study in the Swedish study)(Trollfors et al 1995)

	No. of Cases		
	DTaP Vaccine ($n = 1670$)	DT Vaccine ($n = 1665$)	Vaccine Efficacy [95% CI]
≥21 days of cough			
WHO definition	96	245	63 [52–71]
Göteberg confirmed	77	241	69 [60–77]
Göteberg confirmed + probable	99	252	62 [52–71]
≥21 days of paroxysmal cough			
WHO definition	72	240	71 [63–78]
Göteberg confirmed	58	236	77 [69–83]
Göteberg confirmed + probable	75	246	71 [62–78]
≥7 days of cough			
WHO definition	121	251	54[43–63]
Göteberg confirmed	98	244	62 [51–70]
Göteberg confirmed + probable	125	258	54 [42–63]

agreed to participate. Of these 1724 and 1726 were randomly assigned to DTaP and DT toxoids. There were 817 recipients of DTP toxoids and 850 recipients of DT toxoids with one or more older siblings.

The three vaccine doses were administered intramuscularly at 3, 5, and 12 months. First vaccinations occurred between September 1991 and September 1992, third vaccinations between May 1992 and July 1993. There were 52 children withdrawn from the study for various reasons. Coughing episodes between the first vaccination and July 24, 1994 were included in the study analysis. The surveillance period for each child was divided in two parts. The first part was between the first vaccination until 29 days after the third during which time the children were considered to be incompletely vaccinated. The second part began at the end of the first part for each child and lasted until July 24, 1994. Parents were asked to monitor adverse events for seven days, after which they were interviewed. They were contacted once a month by telephone for further surveillance of adverse events.

Parents were asked to contact the study nurse if anyone in the family coughed for seven or more days. Biological confirmation was done by culture or PCR of a nasopharyngeal sample and serology. Follow-up of each case continued for at least 60 days or until the cough ended. PCR was able to distinguish pertussis from parapertussis. The case definitions were similar to those of the Niakhar study, but the Göteberg group had their own classifications in addition to the WHO criteria. Essentially the Göteberg group allowed that household contacts for the epilink could be confirmed either by culture or serology, whereas the WHO definition allows only culture. The Göteberg group also distinguished two levels of biological evidence. Confirmed cases required two confirmation criteria, and probable cases required

only one (Trollfors et al 1995). To measure immunogenicity, serum was obtained from 3361 children at least four weeks after the third vaccination. IgG antibodies against pertussis toxin and toxin-neutralizing antibodies were measured.

Vaccine efficacy, $VE_{S.IR}$, was based on the ratio of the incidence rates in the DTaP compared to the DT group. Confidence intervals were estimated by an exact calculation based on the conditional binomial distribution that follows from the assumption of a Poisson distribution for cases in each group (Clopper and Pearson 1934). Proportions were compared using a two-sided Fisher's exact test.

Of the 2037 coughing episodes lasting at least seven days, 465 (160 in the DTaP-toxoids group and 305 in the DT-toxoids group) met the criteria for confirmed or probable pertussis, including 368 that met the WHO definition. Another 14 children had clinical pertussis without laboratory confirmation. Thirty days after the third vaccination, 1670 and 1665 recipients of the DTaP and DT toxoids were still at risk for pertussis. The incidence of pertussis according to the WHO definition was 2.96 cases per 100 person-years among the DTaP toxoids recipients and 10.32 cases per 100 person-years in the DT toxoids recipients. The efficacy of the pertussis vaccine was 71% (Table 6.3).

As in the Niakhar pertussis study, the number of cases and the vaccine efficacy estimates vary with the case definitions. The estimates using $\geq$21 days of paroxysmal cough had the highest estimates, $\geq$21 days of any cough the middle estimates, and $\geq$7 days of cough the lowest estimates, reflecting the differing specificity of the case definition. Depending on the case definition used, over 15% of the children in the DT toxoids group developed pertussis during the trial. Although not discussed in detail in this book, the pertussis-toxin testing for defining a case had much lower sensitivity in recipients of DTP toxoids than in recipients of DT toxoids because the DTP-toxoid recipients already had high values for IgG antibodies against pertussis toxin in the acute-phase serum samples. Cultures and PCR were also less sensitive in vaccinated children. A study to estimate the indirect effects of vaccination was nested in this trial (Sections 10.2.5 and 12.5.1)

The acellular pertussis component of the vaccine in the Trollfors et al (1995) study had just the pertussis toxoid. Further acellular vaccine candidates were developed that contained additional antigens. Pertussis toxoid (PT) was included. Other antigens included were filamentous hemagglutinin (FHA), pertactin (PRN), and fimbriae types 2 and 3 (FIM). Two coordinated trials were conducted in Sweden as part of an international effort. Trial I was conducted during the years 1992–1995 (Gustafsson et al 1996). Two acellular pertussis vaccines, one whole cell pertussis vaccine, and one placebo were used. The placebo group ($n = 2574$) received diphtheria and tetanus toxoid (DT). The second group ($n = 2566$) received DTaP2 with two antigens, PT and FHA. The third group ($n = 2587$) received DTaP5 with PT, FHA, PRN, and the two FIM antigens. The fourth group received DTwP. A study to evaluate immunological surrogates of protection after household exposure to pertussis was nested in the primary efficacy study Trial I (Storsaeter et al 1998) (Section 15.3.2).

Trial II was conducted during the years 1993–1996 (Olin et al 1997) with no placebo group. The DTaP5 contained higher amounts of PT and FHA than the

DTaP5 of Trial I. The DTaP2 had the same composition as in Trial I. The source for DTwP in Trial II was different from that in Trial I.

6.4.3 Absolute efficacy of live attenuated influenza vaccine in children

Belshe et al (1998) conducted a randomized, double-blind placebo-controlled multicenter study of the efficacy of live attenuated cold-adapted trivalent influenza virus vaccine in children. Healthy children who were 15 to 71 months of age at time of recruitment and who had no contraindication were enrolled. The vaccine contained that year's recommended strains of influenza A (H1N1), influenza A (H3N2), and influenza B. Children were randomized 2:1 to receive vaccine or placebo. Vaccine was given either as a one- or two-dose regimen, with some of the sites using one or the other. Vaccine and placebo were administered with an intranasal spray applicator.

To evaluate side effects of vaccination, parents were asked to monitor and to record certain symptoms for 10 days after vaccination. They were given a thermometer to measure the temperature. Serious adverse events were followed throughout the trial. Strain-specific immunogenicity of the vaccine was measured in a substudy of 203 participants, approximately the first 21 children recruited at each site. The serum samples were assayed for presence of hemagglutination-inhibiting antibodies to the three viral strains contained in the vaccine.

The primary efficacy endpoint was the first episode of culture-confirmed influenza for subjects who became ill 28 days or more after the receipt of the first dose of vaccine or placebo or at any time after the second dose during the influenza season. Parents were contacted by telephone every two to three weeks until the beginning of an influenza outbreak in their community. Then weekly contact was made to remind the parents to report any relevant symptoms as soon as possible. Study staff attempted to collect specimens for culture for influenza virus confirmation within four days of the onset of symptoms. A case of influenza was defined as any illness detected by active surveillance that was associated with a positive culture for wild-type influenza virus.

The analysis was based on the $\mathrm{VE}_{S,CI}(T)$, using the observed proportions of cases in vaccine recipients and placebo recipients. Koopman's (1984) method for the ratio of two binomials was used to estimate 95% confidence intervals. A logistic generalized estimating equation (Liang and Zeger 1986) with an exchangeable covariance matrix was used to rule out the possibility of an effect within families on the results, because more than half the children in the study were in households with at least two children in the household.

Enrollment began in August 1996 with 1314 children enrolled in the two-dose cohort and 288 in the one-dose cohort. Surveillance ended April 1997 at the end of the influenza outbreaks at the study sites. Among children in the immunogenicity substudy, younger children were more likely to be seronegative before entering the

Table 6.4 Efficacy, $VE_{S,CI}(T)$, of one or two doses of live attenuated, cold-adapted influenza virus vaccine for the prevention of culture-confirmed influenza (Belshe et al 1998)

Influenza Type	Assigned to One Dose			Assigned to Two Doses Who Received Two Doses			All Participants		
	No. of Cases		Efficacy	No. of Cases		Efficacy	No. of Cases		Efficacy
	Vaccine ($n = 189$)	Placebo ($n = 99$)	[95% CI]	Vaccine ($n = 849$)	Placebo ($n = 410$)	[95% CI]	Vaccine ($n = 1070$)	Placebo ($n = 532$)	[95% CI]
A(H3N2)	2	8	87 [47–97]	4	49	96 [90–99]	7	64	95 [88–97]
B	1	6	91 [46–99]	6	31	91 [78–96]	7	37	91 [79–96]
Any	3	14	89 [65–96]	10	74	94 [88–97]	14	95	93 [88–96]

study than older children. Only 29% of children one or two years of age had antibodies to influenza A (H3N2) compared with 70% of children three years of age or older. Pre-existing antibody to influenza in an influenza vaccine study is considered an important potential confounder. Of the 3009 illnesses in the study subjects, 71 cases of influenza A (H3N2) and 44 cases of influenza B were confirmed. No cases of wild-type influenza A (H1N1) were identified in the study participants or the communities at large during the 1996–1997 influenza season. Table 6.4 shows the results. Vaccination was quite effective against culture-confirmed influenza. Although the data are not presented here, the spectrum of illness in the vaccinated children who developed influenza was milder than that in unvaccinated children.

In influenza vaccine studies for vaccines directed against annual influenza, there is generally an attempt to get all of the participants vaccinated before the beginning of the influenza season. Then given the short duration of the season, influenza vaccine studies can often use an analysis based on the simple cumulative incidence or attack rates. The trial continued beyond the first year. Longini et al (2000) analyzed the first and second year of the trial, allowing for site-specific attack rates. There was some evidence that study sites with high attack rates the first year had lower attack rates the second year and vice versa, suggesting a possible herd immunity effect.

6.4.4 Live attenuated influenza vaccine in adults without biological confirmation

A randomized, double-blind, placebo-controlled trial of live attenuated trivalent influenza virus vaccine in healthy adults was conducted from September 1997 through March 1998 in 13 centers across the United States (Nichol et al 1999). Three of the main outcome measures were episodes of febrile illness, severe febrile illness, and febrile upper respiratory tract illness. Cultures were not performed for confirmation of influenza illness and culture-confirmed influenza was not an outcome in contrast to the Belshe et al (1998) study in young children. Nichol et al (1999) called this an effectiveness study, not an efficacy study. Participants were enrolled mid-September

Table 6.5 Efficacy (effectiveness) of live attenuated, cold-adapted influenza virus vaccine for the prevention of some clinical outcomes (not culture-confirmed influenza) in adults (Nichol et al 1999)

Outcome	Vaccine Group		Placebo Group			
	Total Episodes No. (n=2833)	Rate per 1000 Persons per 7-Week Outbreak	Total Episodes No. (n=1420)	Rate per 1000 Persons per 7-Week Outbreak	Reduction in Rates, % [95% CI]	*p*-value
Febrile illness	406	151.3	225	168.1	10.0 [−2.1–20.7]	.10
Severe febrile illness	298	111.0	183	136.7	18.8 [7.4–28.8]	.002
Febrile upper resp tract illness	248	92.4	162	121.0	23.6 [12.7–33.2]	<.001

to mid-November 1997. Recruitment strategies varied across sites. Persons were eligible if they were 18 to 64 years old, worked at least 30 hours per week outside the home, had health insurance, and were reachable by telephone. There were the usual exclusion criteria. The vaccine contained the three viruses corresponding to those recommended for the 1997–1998 influenza season in the United States. Vaccines were administered intranasally between September 18 and November 15, 1997.

Participants were randomized 2:1 to receive the vaccine or placebo in the fall of 1997. A total of 3041 adults received vaccine and 1520 received placebo. Reactogenicity and safety were assessed by asking each participant to keep a record of daily symptoms on the evening of vaccination and seven days afterwards. Participants were called at day 28 to identify serious adverse events. Assessment of any serious adverse events continued to the end of the study. Influenza virus surveillance is conducted in many places across the United States. The surveillance identifies the influenza season and the strains of circulating wild-type virus. Nichol et al (1999) identified two influenza outbreak periods. The first was the site-specific peak outbreak, using the modal week at each site to begin an algorithm that identified the weeks in which at least 80% of the positive influenza isolates for the season were included. The total outbreak period was identified by a panel of experts from the surveillance information from all of the sites. The motivation for choosing the site-specific outbreak period was that the identified cases would have a higher probability of being influenza.

Bivariate comparisons for the proportions of subjects experiencing study outcomes were conducted using the Cochran–Mantel–Haenszel test controlling for site. Generalized linear models were used to calculate the variance of the event rates.

At the different sites, peak outbreak periods lasted from 4 to 12 weeks. The surveillance cultures revealed that nearly all of the isolates that year were influenza A (H3N2), 80% of which were a drifted variant of the vaccine strain, so that the vaccine was not well matched to a large portion of the circulating viruses.

Table 6.5 contains the results for three of the main outcome measures. For the most specific case definition of febrile upper respiratory tract illness, the efficacy of the vaccine is 23.6% (95% CI, 12.7–33.2), much lower than in the Belshe et al (1998) study the previous year in young children (Table 6.4). There are three possible explanations for the lower efficacy. First, the circulating strain was a drifted variant of the vaccine strain. Second, some investigators believe that adults do not respond to the intranasal live attenuated vaccine as well as children. Third, the case definition is not confirmed influenza, so that many of the illnesses captured in the analysis are likely not influenza, causing the efficacy estimates to be much lower than efficacy estimates of culture-confirmed influenza. This latter reason certainly played an important role. In Chapter 8 we show how to estimate vaccine efficacy for a biologically confirmed outcome when only a small subsample of the nonspecific cases are biologically confirmed.

6.4.5 Relative efficacy of live and killed influenza vaccine in young children

Soon after universal vaccination of children 6 to 59 months of age was recommended by the U.S. advisory bodies, a double-blind randomized trial in infants and young children to compare live attenuated influenza vaccine with inactivated influenza vaccine was conducted (Belshe et al 2007).

The study was conducted at 249 sites in 16 countries in the United States, Europe, the middle East, and Asia. The sites were physicians' offices and primary care clinics. Children were randomly assigned on a 1:1 basis to receive one of the two vaccines. Subjects were stratified in the randomization to age on receipt of the first dose, presence or absence of previous influenza vaccination, presence or absence of wheezing, and country of residence. The usual exclusion criteria applied. Children with mild or moderate asthma or wheezing more than 42 days before the trial were included. Children not previously vaccinated for influenza received two doses of the assigned study vaccine. To preserve blinding, children assigned the intranasal live attenuated vaccine received an intramuscular injection of salt solution, and analogously for children assigned the intramuscular killed vaccine.

Parents recorded local and systemic reactions until 42 days after vaccination. Medically significant events were collected throughout until the end of the surveillance period, May 31, 2005. Study staff contacted the parents every 7 to 10 days during the surveillance period. Nasal swabs for viral cultures were obtained either at the child's home or at the study site. The study was powered assuming a 3.0% attack rate in children receiving killed vaccine and a 1.8% attack rate in the children receiving live attenuated vaccine, for a relative efficacy of 40%. Assuming that 90% of the children would be able to be included in the per-protocol analysis, 8500 children would be needed for 90% power to demonstrate superiority of the live attenuated to the inactivated vaccine. The primary endpoint was the relative efficacy in preventing culture-confirmed influenza-like illness caused by well-matched influenza strains.

Table 6.6 Relative reduction in attack rate with live attenuated, cold-adapted influenza virus vaccine compared to inactivated vaccine regardless of match for the prevention of culture-confirmed influenza in infants and young children (Belshe et al 2007)

Virus	Live Attenuated Vaccine (n=3916)		Inactivated Vaccine (n=3936)		Reduction in Attack Rate
	Cases No.	Attack rate %	Cases No.	Attack Rate %	With Live Vaccine % [95% CI]
All	153	3.9	338	8.6	54.9 [45.4–62.9]
A/H1N1	3	0.1	27	0.7	89.2 [67.7–97.4]
A/H3N2	37	0.9	178	4.5	79.2 [70.6–85.7]
B	115	2.9	136	3.5	16.1 [−7.7–34.7]

The definition of influenza-like illness was an oral temperature of 37.8°C or higher or the equivalent in the presence of cough, sore throat, or runny nose or nasal congestion occurring on the same or consecutive days. Secondary endpoints included relative efficacy against mismatched influenza viruses and all influenza viruses, as well as several other clinical outcomes, such as otitis media.

From October 20 to October 29, 2004, a total of 8475 children were enrolled. Of these, 7852 were included in the per-protocol analysis. Table 6.6 shows the overall number of cases regardless of match of the vaccine with the circulating strains. The paper presents analysis by well matched vaccine, well-matched by age group, well matched by previous vaccination status, and not well matched. In this trial, of the 3936 children who received inactivated vaccine, 338 developed culture-confirmed cases of influenza. Of the 3916 children who received live attenuated vaccine, 153 cases developed. Relative reduction in attack rate by the live vaccine compared to the killed vaccine was 54.9% (95% CI 45.4–62.9).

6.4.6 Oral cholera vaccines in Bangladesh

Interest in oral cholera vaccines developed because parenteral vaccination had not been very successful. Cholera is a disease in the intestine, so it seemed that local mucosal immunity stimulated by an oral vaccine might be better. A randomized, double-blind trial of two oral killed cholera vaccines and one placebo arm was conducted in the Matlab field studies area of the International Centre for Diarrheal Research, Bangladesh (ICDDR,B) (Clemens et al 1986). The oral vaccines consisted of killed cholera whole cells (WC) either with or without the B subunit (BS) component of cholera toxin. The placebo arm received a heat-inactivated *E.coli* K12 strain.

Potentially eligible subjects for the trial were the 124,035 persons aged 2 to 15 years and females aged over 15 years residing in the vaccine trial area at the onset of vaccination. These are the groups at highest risk for cholera in Matlab. After exclusion criteria, 89,596 persons took at least one dose of vaccine or placebo. A census

Table 6.7 Occurrence of cholera and $VE_{S,CI}(T)$ during the first year of follow-up after the third dose among participants who ingested three complete doses of the vaccine or placebo assigned (Clemens et al 1988)

	Group				
Outcome	BS-WC No.	VE %	WC No.	VE %	K12 No.
Cholera	41	62	52	53	110
No cholera	20,664		20,691		20,727
Total	20,705		20,743		20,837

of the vaccine trial population was conducted three months prior to vaccination. Persons were randomized in the census to one of the three groups before teams went to the field. Vaccination occurred in three six-week rounds starting in January, 1985, with a short one-week round in May, 1985. Vaccines and placebo were delivered by 69 vaccination teams who were assigned to particular villages and visited people in their homes. The estimated fraction of the oral dose swallowed was recorded. Physicians in the trial area were stationed during vaccination to manage side effects.

Surveillance for diarrhea was maintained at the three diarrheal treatment centers serving the Matlab population. Stool samples or rectal swabs were processed to identify *V cholerae* 01, and to determine the biotype (El Tor or classical) and serotype of each isolate. To be considered fully vaccinated, a person needed to have three doses, and have swallowed all of the first dose and at least 3/4 of the second and third doses. Later follow-up analyses focused on those participants who had completely ingested all three doses (Clemens et al 1988; Clemens 1990). The case definition was that the participant presented for treatment of diarrhea whose onset was $\geq$14 days after receipt of the third dose, had various diarrheal symptoms not detailed here, *V. cholerae* was isolated, and a field check at the person's home confirmed that the person had indeed sought treatment on the specified date.

The vaccine efficacy measure after one year of follow-up was based on the proportion of vaccinees compared to the proportion of controls becoming ill with cholera, $VE_{S,CI}(T)$ (Clemens et al 1988). Table 6.7 presents the analysis of one year of follow-up. Cases were those presenting with onset between 14 and 365 days after the third dose. In this analysis, only those who ingested three complete doses were included. Of those initially enrolled in the study, 62,285 participants took three complete doses of either placebo, whole cell, or B-subunit whole cell vaccine, with 20,837, 20,743, and 20,750 in each group. The group reported one-sided confidence intervals, which are not included in Table 6.7 (see Problem 6.1 and Table 7.3). In subsequent years of follow-up, the efficacy of the vaccines appeared to wane. In Section 7.3 we present a method to analyze vaccine efficacy that wanes over time using the example of the cholera vaccine trial.

6.4.7 Pneumococcal conjugate vaccine in California

A randomized, double-blind trial of a heptavalent pneumococcal vaccine was conducted at 23 medical centers within Northern California Kaiser Permanente (NCKP), a health maintenance organization (Black et al 2000). Healthy infants were randomized 1:1 to receive either heptavalent pneumococcal conjugate or the meningococcus type C conjugate vaccine at 2, 4, 6, and 12 to 15 months of age. Infants with specific risk factors were excluded. The heptavalent vaccine contained saccharides of the serotypes 4, 9V, 14, 18C, 19F, 23F, and 6B conjugated to a protein carrier made of nontoxic mutant diphtheria toxin. At that time, the seven serotypes were responsible for 83% of invasive disease in children younger than 4 years of age. The control meningococcal conjugate vaccine had the same carrier.

The primary endpoint was invasive pneumococcal disease caused by the vaccine serotypes. Secondary endpoints included otitis media. The outcome pneumonia was reported separately from the primary analysis. Active surveillance for cases in the study population was conducted using automated clinical and laboratory databases of the NCKP system. Invasive pneumococcal disease was defined as a positive culture of *Streptococcus pneumoniae* from a normally sterile body fluid (blood, spinal fluid) obtained from a child presenting with an acute illness compatible with pneumococcal illness.

Between October 1995 and August 1998, 37,868 children were enrolled into the trial. Of the 18,927 children who received at least one dose of pneumococcal conjugate, 17,174 received at least two doses, 15,565 received at least three doses, and 10,940 received at least four doses. Of the 18,941 children who received at least one dose of meningococcal conjugate, 17,196 received at least two doses, 15,536 received at least three doses, and 10,995 received at least four doses.

In this trial, protective efficacy was estimated by 1 minus the ratio of the number of cases of invasive disease in the pneumoccal vaccine arm compared to the meningococcal arm. In other words, the computation does not use the denominators. Efficacy was evaluated with the binomial test of the null hypothesis that the vaccine has no efficacy for the seven serotypes. The analysis incorporated a sequential design. An interim analysis had been planned when 17 cases had occurred. The null hypothesis was to be rejected if the case split was 15:2 or more favorable, $p = 0.0023$, with a final evaluation planned when 26 cases had occurred and an overall two-tailed p value of <0.05. Exact binomial confidence intervals were calculated by the Clopper–Pearson (1934) method. An intent-to-treat analysis included all invasive disease caused by a pneumococcal serotype regardless of number of doses completed. Safety of the vaccine was assessed by telephone follow-up on subsets of the study population, one receiving DTwP, one receiving DTaP. The computerized utilization data of the NCKP was also used to compare rates of events in the two groups. Immunogenicity of the conjugate vaccine was evaluated in a subset of children receiving DTwP concurrently and in a subset given DTaP in the first year of life. Serum IgG to the seven serotypes was measured using ELISA from samples collected before the first vaccination and one month after the third dose.

Table 6.8 Efficacy of heptavalent pneumococcal vaccine against invasive pneumococcal disease results as of April 20,1999 (Black et al 2000)

Analysis for Serotypes Contained in the Vaccine	Cases Split Control: Pneumococcal Vaccine Groups	Efficacy % [95% CI]	*p*-value
Per protocol fully vaccinated	39:1	97.4 [82.7–99.9]	<0.001
Intent to treat	49:3	93.9 [79.6–98.5]	<0.001
Partially vaccinated only	7:1	85.7 [0–100]	0.05
All cases regardless of serotype	55:6	89.1 [73.7–95.8]	<0.001

At the interim analysis, all 17 of the cases of invasive disease in fully vaccinated children were in the control group. At the interim intent-to-treat analysis of children receiving at least one dose, all 22 cases were in the control group. The Study Advisory Group recommended termination of the trial at the interim analysis because of the high efficacy. Enrollment was discontinued at the end of August 1998. Blinded follow-up and per-protocol vaccination of the two groups continued until April 20,1999. After that, all children in the control group were offered pneumococcal conjugate vaccine. The vaccine was highly efficacious against invasive pneumococcal disease (Table 6.8). During the trial, concern grew that there would not be enough events for the definitive analysis. This motivated the design and implementation of the grouprandomized study to estimate the total effects of using the pneumococcal vaccine (Section 13.4.2).

6.5 Report of a Study

In the preceding examples we have not included every aspect of the report of the studies. A report should tell the type of study, whether randomized, cohort, or case-control. The entities that reviewed the study protocol should be listed. These could include local institutional review boards, regulatory bodies, such as the U.S. Food and Drug Administration, medical products committees, and ethics boards. Details of the vaccines and placebos, their manufacturers, the lots, and any other relevant aspect such as storage should be included. Details of the route and schedule for administering the vaccines are needed. The study description should include the usual person, time, and place. The study population, the eligibility for inclusion, the dates for eligibility, exclusion criteria, how cases were ascertained, the case definition(s), the follow-up period, and where the study took place all should be included. The surveillance for side effects or adverse events, the laboratory methods if any for biological confirmation of cases, reasons for loss to follow-up, and immunogenicity tests, should be described. The statistical analysis and possibly how the sample size was chosen should be described. The results usually include a descriptive comparison of the groups on important potential confounders. Reports of randomized

controlled trials can follow the Consolidated Standards of Reporting Trials (CONSORT) Statement (Moher et al 2001; Altman et al 2001).

6.6 Reduction in Burden of Illness

Most of the studies of VE_S presented in this chapter have a case definition that is a 0,1 dichotomous outcome. Although several different case definitions, some more and some less severe, may be considered in separate analyses, they are all scored 0,1 in any given analysis. Chang et al (1994) suggested a measure of efficacy that takes into account both the incidence of disease and severity. A severity score is assigned to each incident case, with 0 assigned to noncases. Then the total is summed over all cases to have a burden of illness score. When the severity score for each case is one, the burden-of-illness score reduces to the vaccine efficacy based on the number of cases in the vaccinated compared with the unvaccinated group. When different cases have different severity scores, the burden-of-illness score for a group is a weighted sum of all of the cases in the group, where the severity scores serve as the weights. The burden-of-illness score divided by the number of subjects randomized to the group yields the burden-of-illness per randomized participant. The difference between the mean burden-of-illness in the two groups, or the relative difference is a measure of the net reduction in morbidity per participant. The reduction in burden of illness differs from the VE_P measures in that the denominator is still the susceptible people, and the first outcome post-randomization is illness, which is given a score. A number of vaccine studies have developed severity scores (Section 9.2.1). In a rotavirus vaccine study, the severity of each case of diarrhea was given a severity score between 0 and 20 (Ruuska and Vesikari 1990).

Let N_0 and N_1 be the number randomized to vaccine and control, and c_0 and c_1 the number of cases in the vaccine and control arms. The severity scores for the cases are $S_{01},\ldots,S_{0n_0}$ and $S_{11},\ldots,S_{1n_1}$ in the two groups with means μ_0, μ_1 and variances σ_0^2, σ_1^2. One design option is that the trial runs for a fixed time, after which it is stopped and analyzed. A second option is that the trial is stopped after a number of total cases c, where $c = c_0 + c_1$. If λ_0 and λ_1 are the hazards of disease in the two groups, then the expected number of cases in the two groups is $\lambda_0 N_0 t$ and $\lambda_1 N_1 t$, where t is the duration of follow-up. The number of cases in the control group, c_0, has a binomial distribution $\text{Binom}(c, p_0)$, where $p_0 = \lambda_0 N_0 t/(\lambda_0 N_0 t + \lambda_1 N_1 t)$, and $p_1 = 1 - p_0$. In the design with fixed time, the null hypothesis is that $\mu_0 = \mu_1$ and $p_0 = p_1$. In the design with fixed number of events, the null hypothesis is that $\mu_0 = \mu_1$ and $\lambda_0 = \lambda_1$. A test statistic T for both models is the difference in the mean burden of illness scores per participant:

$$T = \frac{1}{N_0}\sum_{i=1}^{n_0} S_{0i} - \frac{1}{N_1}\sum_{i=1}^{n_1} S_{1i}. \tag{6.16}$$

For both designs, under the null hypothesis, μ_0 and μ_1 are estimated by

$$\bar{x} = \left(\sum_{i=1}^{n_0} S_{0i} + \sum_{i=1}^{n_1} S_{1i} \right) / (n_0 + n_1) = (n_0 \bar{s}_0 + n_1 \bar{s}_1)/(n_0 + n_1). \tag{6.17}$$

The variances $j = 0, 1$ are estimated by

$$s_j^2 = \left(\sum_{i=1}^{n_j} (S_{ji} - \bar{S}_j)^2 \right) / (n_j - 1). \tag{6.18}$$

In the fixed time design, $\hat{p} = (n_0 + n_1)/(N_0 + N_1)$ estimates both p_0 and p_1. In the fixed number of events design, p_0 is estimated by $N_0/(N_0 + N_1)$, and p_1 by 1 minus the estimate of p_0. The observed standard test statistics are obtained from

$$\widehat{V}_H(T) = [\bar{x}^2 \hat{p}(1-\hat{p})/(1/N_0 + 1/N_1) + \hat{p}(s_0^2/N_0 + s_1^2/N_1)]$$
$$\widehat{V}_H(T|n) = c[\bar{x}^2/N_0 N_1 + (s_0^2/N_0 + s_1^2/N_1)/(N_0 + N_1)].$$

The two-sided rejection region of the null hypothesis for the fixed time design is $|T/\sqrt{\widehat{V}_H(T)}| > z_{\alpha/2}$ and for the fixed number of events design is $|T/\sqrt{\widehat{V}_H(T|n)}| > z_{\alpha/2}$. Chang et al (1994) also present a method to calculate sample size. Because the scores combine incidence with severity per case, one might think that the burden-of-illness scores can provide a more comprehensive measure of overall efficacy than would a separate analysis based simply on either incident cases, VE_S, or the per-case severity, VE_P with a continuous outcome (Chapter 9), alone. However, because there may be a large number of zeros in each group, the test can have poor power.

Mehrotra et al (2006) compared eight methods for a dual endpoint evaluation of efficacy in a proof-of-concept trial, including that of Chang et al (1994). The motivation for the comparison was the design of the first trial of an HIV vaccine based on cell-mediated immunity. The vaccine was expected to have very low efficacy against infection, but it was hoped that it would reduce viral load as a surrogate for progression to disease. The question was whether it was better to test the composite null hypothesis of no vaccine effect on either the incidence of HIV infection or the viral load setpoint among those who become infected relative to the placebo using just a single composite test or using two separate tests, one for the infection endpoint and one for viral load endpoint. They found that combining separate tests for the infection and viral load endpoints is generally more powerful than the unconditional burden-of-illness test of Chang et al (1994), especially at low or zero VE_S. At $VE_S = 0.60$ or higher, all methods and combinations of methods performed comparably. They recommended using either the unweighted Simes' or Fisher's combination test for the trial.

One of the problems in vaccine studies is that usually most of the participants do not become infected. Follmann et al (2009) took a different approach from that of Chang et al (1994) by introducing chop-lump Wilcoxon and t-tests. The approach again assigns a score S to each participant, 0 for uninfected participants, and a measure $S > 0$ of the post-infection outcome such as severity or parasite density in the infected participants. When the number of participants in each group is equal,

the chop-lump test first removes an equal number of zeros from both groups, then performs the test on the remaining S scores, most of which are greater than 0. A permutation approach then provides a null distribution. The chop-lump Wilcoxon test is shown to be always more powerful than the usual Wilcoxon test when the true infection rates in the vaccine and the control group are the same. The R package `choplump` is available at http://cran.r-project.org/.

Problems

6.1. (a) Cholera study: compute one-sided and two-sided 95% confidence intervals for VE in Table 6.7. (b) Compare the results. (c) Why are two-sided confidence intervals generally recommended?

6.2. A randomized study of an influenza vaccine was conducted with 3000 children each in the vaccine arm and the control arm. There were 350 biologically confirmed cases in the control arm and 53 cases in the vaccine arm by the end of the influenza season. Compute the estimate of $\mathrm{VE}_{S,CI}(T)$ and the 95% confidence limits on the estimate.

6.3. (a) In an observational study in a cohort, some of whom are vaccinated and some not, how might the exposure to infection differ in the two groups?
(b) Would differing exposure to infection be a confounder in the study? How might it influence the vaccine efficacy estimates using $\mathrm{VE}_{S,CI}(T)$ or $\mathrm{VE}_{S,IR}$? Write out $\mathrm{VE}_{S,CI}(T)$ and $\mathrm{VE}_{S,IR}$ using the dependent happening expression (2.7) to explain your response.
(c) How might you ascertain differences in exposure to infection or control for it in the analysis?
(d) How would this vary for different infectious diseases?

6.4. Discuss how and why the vaccine efficacy estimates in Table 6.3 change with the changing case definition.

6.5. (a) Consider designing a relative efficacy trial of a live, attenuated influenza virus vaccine with a killed influenza virus vaccine. Assume a 5.0% attack rate in the children receiving killed vaccine and a 2.5% attack rate in the children receiving live, attenuated influenza virus vaccine. How large a sample size would be needed in each arm for 90% power with $\alpha = 0.5$ on a two-sided test? (b) Assume now attack rates of 1.0% and 0.05% in the two arms. What sample size would be needed in each arm to achieve the same power and α level?

6.6. (a) Explain the main difference between the approach of Chang et al (1994) in testing for differences in burden-of-illness in the vaccine and control groups and the chop-lump test of Follmann et al (2009).