15

Surrogates of Protection

15.1 Replacing clinical outcomes

A holy grail of vaccine research is to identify a vaccine-induced immune response that predicts protection from infection and disease. If a measurable immune response to vaccination were available that were predictive of protection from infection and disease, it would help to avoid new large trials and facilitate getting new products and formulations approved. An immunological surrogate of protection could reduce the sample size or shorten the duration of a trial. Thus, identifying a a good immunological surrogate of protection could make a trial be much less expensive or indeed be feasible. If a good vaccine is already licensed and recommended, a trial with a new vaccine compared to placebo would be unethical. When both vaccines are highly efficacious or the clinical outcome of interest is rare, a relative efficacy trial comparing the two vaccines would be prohibitively large.

Because the interest is in evaluating new vaccine candidates in different populations, the primary goal is to predict how well the vaccine will do in new situations. Another use of immunological surrogates of protection is in designing vaccines for infectious agents such as possible future emerging pathogens such as pandemic influenza or anthrax in which clinical outcome data are not available. These latter two types of studies are sometimes called bridge studies.

Much of this book has considered the effects of vaccination on clinical outcomes and on transmission measured by clinical outcomes. The era of using clinical outcomes in most primary vaccine efficacy trials may slowly be coming to an end, though clinical outcomes will still be useful in observational studies. In 1993, a Hib conjugate vaccine was approved for licensure based on immunological data (Frasch 1994) following the licensure of two others based on phase III efficacy trials (Black et al 1992; Santosham et al 1991). Meningococcal C conjugate vaccines were licensed on the basis of serological correlates of protection without Phase III efficacy data (Andrews et al 2003). Identifying immunological correlates of protection is one of original topics of the Gates

Grand Challenges. In this chapter, we present methods to assess correlates and surrogates of vaccine protection. The main focus is on immunological surrogates of protection, but we also consider briefly on the use of carriage as an endpoint in pneumococcal vaccine studies, the subject of ongoing research.

15.2 Biological versus statistical issues

In other fields, considerable interest developed in what were called surrogate endpoints as replacements for a primary clinical endpoint. Over the years, much methodological discussion has revolved on what constitutes a close relationship between the true endpoint and the potential surrogate endpoint. In the vaccine literature, traditionally the term correlate of protection has been used to describe the relation of a vaccine-induced response to the clinical infection or disease outcome. Several different concepts were covered by the term correlate of protection. In vaccine studies, part of the problem is biological and part of the problem is methodological.

The biological problem has several different aspects to it. The main scientific problem to identify a candidate immunological measure or several measures likely associated with clinical protection. A statistical approach cannot validate an immunological measure as related to protection if a candidate has not been identified. Regarding a candidate immunological measure, several aspects need to be delineated. The time of the assay after vaccination, and in the case of multiple doses, the timing after which dose, needs to be decided. The choice of assay can be important. Some assays are more sensitive than others, resulting in different response profiles. The type of antibody measure can play a role. Assays can measure either the antibody concentration, the antibody avidity, or the concentration of functional antibodies. The avidity measures the total strength of the binding of the antibody with the antigen. The avidity can be high for bacteria that can have multiple identical sites. Antibodies with higher avidity can eliminate an antigen at lower concentrations than antibodies with low avidity. Maturation of antibody avidity is a sign of the presence of immunological memory. Functional antibodies may be demonstrate bactericidal activity in assays using whole blood. Another issue is whether an assay measures short-term protection or long-term protection from immunological memory.

As an example, the serum bactericidal assay (SBA) titer was established by Goldschneider et al (1969) as a correlate of protection for meningococcal C disease using a human complement assay. More recently, however, the rabbit complement assay has been recommended. Since the two assays have different sensitivities, the protective titers needed to be re-evaluated (Andrews et al 2003). Serological correlates of protection for meningococal serogroup C can also be measured using avidity, which may be indicative of successful priming of the memory responses by vaccination. The SBA titer may be a correlate of short-term protection and the avidity, as a measure of immunological memory,

may be a measure of long-term protection (Balmer and Borrow 2004). These issues are relevant for the planned licensing of the meningococcal A vaccine using immunological measures alone.

The methodological problem is to validate the identified potential correlates and surrogates of protection. There are two distinct but related problems. One is to identify immunological markers predictive of protection. The second is to identify immunological markers predictive of **vaccine-induced** protection. The relation of these two questions are discussed in this chapter. The correlates of protection may be based on individual measurements or population level measures (Siber 1997).

One of the problems in evaluating correlates of protection is the not everyone in the group under observation is exposed to infection. Thus, a person might not develop disease because of not being exposed, not necessarily because of being protected. A simple general approach assumes the probability of disease is the product of the probability of disease if not protected and the probability of not being protected:

$$\Pr[\text{disease}] = \Pr[\text{disease}|\text{not protected}] \times \Pr[\text{not protected}]. \tag{15.1}$$

In a study, the probability of disease can be estimated by the attack rate or cumulative incidence. Thus, vaccine efficacy based on the attack rate or cumulative incidence can be written

$$\begin{aligned} \text{VE}_{S,CI} &= 1 - \frac{\Pr[\text{disease (vac)}]}{\Pr[\text{disease (controls)}]} \\ &= 1 - \frac{\Pr[\text{disease}|\text{not protected (vac)}]\Pr[\text{not protected (vac)}]}{\Pr[\text{disease}|\text{not protected (control]}\Pr[\text{not protected (control)}]}. \end{aligned} \tag{15.2}$$

Under the assumption that exposure to infection is equal in the vaccinated and control groups, and that the probability of disease is equal if exposed, the terms for the probability of disease if not protected cancels, leaving

$$\text{VE}_{S,CI} = 1 - \frac{\Pr[\text{not protected (vac)}]}{\Pr[\text{not protected (control)}]}. \tag{15.3}$$

The probability of not being protected can be based on a threshold level of antibody above which everyone is protected. Then the probability of being protected is estimated by the proportion of people with immune response above the threshold. Alternatively, one can estimate the probability of protection as a continuous function of the level of antibody, and then base the $\text{VE}_{S,CI}$ in equation (15.3) on the average probability of being protected in the vaccinated compared with the control group using the estimated individual probabilities of being protected at each antibody level. A special case occurs if everyone is exposed to infection, as in challenge studies. Household exposure

to infection has been used as a natural challenge. The probability of developing disease was modeled directly as a continuous function of the antibody titers (Storsaeter et al 1998), and the probability of disease in the vaccinated and unvaccinated groups of another vaccine study predicted (Kohberger et al 2008). The threshold and regression approaches are presented in Sections 15.3 and 15.4.

All of these models are based on an all-or-none model of vaccine protection, whether based on the threshold or based on a continuous model. In the continuous model, at a given antibody titer, a person is either protected or not with an antibody-specific probability. The model also assumes that the protection conferred by titers produced by natural exposure or vaccination are equivalent.

15.2.1 Background

In a groundbreaking paper, Prentice (1989) proposed four criteria for a biomarker to be a surrogate endpoint for the primary clinical outcome of interest. In the context of vaccines, the four can be stated as

1. Protection is significantly related to the vaccine.
2. The surrogate is significantly related to the vaccine
3. The surrogate is significantly related to the clinical endpoint.
4. The surrogate explains all of the clinical endpoints.

The last criterion can be checked by a statistical regression model that has both the treatment indicator and the value or model for the surrogate in the model. Different approaches can be taken. One could say that if regression coefficient for the treatment indicator is not significantly different from 0, then the criterion is met. In another approach, one could require that the regression coefficient actually be 0, which will generally not happen. Kohberger et al (2008) take an alternative approach to the fourth criteria based on estimation of the proportion of the clinical endpoint explained (PE) by the surrogate (Burzykowski, et al 2005). A large literature on the subject of surrogate endpoints in many contexts followed after the 1989 paper by Prentice.

Frangakis and Rubin (2002) criticized the Prentice approach because it is subject to post-randomization selection bias. In the vaccine context, under the Prentice approach, the risk of the clinical endpoints is compared in individuals with the observed values of the immunological markers. However, we observe only the immunological value and the clinical endpoint that the person has under the actual vaccine assignment. We do not observe the value of the immune marker value that the person would have had under the other vaccine assignment. However, similar to the discussion of VE_P in Chapter 9, comparisons based on the Prentice criteria are subject to a post-randomization selection bias and do not have a causal interpretation. Frangakis and Rubin (2002) call the surrogates evaluated by the Prentice criteria statistical surrogates. Using the framework of potential outcomes in causal inference (Chapter

1.4), they propose a definition of a principal surrogate based on comparison of individuals with the same pair of potential values of the candidate surrogate under the two treatments.

15.3 Thresholds for protection

In a threshold model, let AB(protective) be the level of antibody assumed to be protective. If $\mathrm{VE}_{S,CI}$ based on the clinical outcome is known the antibody level is measured in everyone, and under the assumption that exposure is assumed equal in the two group, then following equations (15.2) and (15.3), we can simply solve for the level of antibody that is protective:

$$\mathrm{VE}_{S,CI} = 1 - \frac{\Pr[Y = 1|\text{vaccinated}]}{\Pr[Y = 1|\text{control}]}$$

$$\mathrm{VE}_{S,CI} = 1 - \frac{\%\text{ of vaccinated with [Ab]} < \text{AB(protective)}}{\%\text{ of controls with [Ab]} < \text{AB(protective)}} \qquad (15.4)$$

in contrast, given a threshold assumed to be protective, we can also predict the vaccine efficacy based on the proportion of people in the vaccinated and unvaccinated and unvaccinated groups who are above that threshold.

Andrews et al (2003) used postlicensure surveillance of meningococcal C to validate the serologic correlates of protection used to license the conjugate vaccine in England. Starting with equation (eq:surro:basic1), they assumed that exposure to infection was the same in the vaccinated and unvaccinated group, and that the protection conferred by titers produced by natural exposure or vaccination are equivalent. They explored the efficacy predicted using equation 15.4 by different cutoff thresholds for protection (Table 15.1). The screening method (Chapter 8.1.4) was used to estimate the observed postlicensure efficacy (direct effectiveness). Cases of confirmed meningitis C infection that occurred in vaccinated and unvaccinated individuals in England from January 2000 to the end of 2001 and coverage levels of vaccination were used for the computation.

In preschool children, there were 27 cases occurred, all in unvaccinated children for an efficacy estimate of 100% (95% CI, 93.3–100%). Coverage levels were not given in the paper. From Table 15.1, the predicted efficacy from titers one month after vaccination is consistent with the observed efficacy at all of the cutoffs except 1:128. However, using titers 7 and 9 months postvaccination, the predicted vaccine efficacy significantly underestimated the observed efficacy in infants and toddlers (preschool children were not included). This suggests that when the postvaccination titers have declined, that immunologic memory and a rapid booster response may be responsible for efficacy, which would better be measured by antibody avidity.

Jódar et al (2003) discuss this approach in the context of multivalent pneumococcal vaccines.

Table 15.1. Predicted vaccine efficacy and 95% CIs estimated for unvaccinated and vaccinated preschool children with titers below the different serum bactericidal assay (SBA) cutoffs one month after vaccination with the meningococcal conjugate vaccine measured by SBA (from Andrews et al 2003).

	% Individual with titers below cutoff		
Cutoff	Vaccinated	Unvaccinated	Predicted % vaccine efficacy (95% CI)
1:4	0.0	90.4	100 (95–100)
1:8	0.0	93.3	100 (95–100)
1:16	2.5	94.3	97 (92–99)
1:32	4.1	95.2	96 (90-98)
1:64	4.9	97.1	95 (89–98)
1:128	9.8	97.6	90 (83–94)

15.4 Regression models for correlates

15.4.1 Logistic regression model

15.4.2 Estimating the other factors

Dunning (2006) points out that models can capture the observed relation between immunological assay and protection from disease at high assay values. At high values very few people develop disease. However, at low assay values, whether a person develops disease could be associated with whether the person as exposed or not. Thus the probability of developing disease in individuals with low assay values could depend on the prevalence of the disease through the dependent happening relation or other factors not associated with the immunological measures. Dunning (2006) proposed a model that separates the effect of the assay values from such factors as level of exposure and disease prevalence. In fitting the data from individual-based measurements with clinical outcome and titers, the model estimates a parameter that represents levels of exposure to infection and other factors not included in the measured immune responses rather than cancelling it out. In the second step, when predicting vaccine efficacy from the estimated regression parameters, the estimated factor is assumed to cancel out.

Assume there are data from$n$ participants, $i = 1, \ldots, n$. Let x_i be the assay value for participant i, and $y_i = 1$ if participant i develops disease, and $y_i = 0$ if not. It is assumed that x is log transformed so that it can have negative values. The model has two main components. The first is the probability $\alpha(x)$ that a person with titer x is protected. The second is the probability ω that a susceptible individual develops disease. The probability $\alpha(x)$ is essentially an all-or-none model of protection where the probability of being completely protected is a function of the immunological assay value. The protected individuals are assumed completely immune from disease, and the $(1 - \alpha(x))$ susceptible individuals are assumed to be homogeneously susceptible.

The probability that an individual develops disease is the product of the probability that the individual is susceptible and the probability that a susceptible individual will develop disease:

$$\Pr(Y_i = 1|X_i = x) = \omega(1 - \alpha(x_i)). \tag{15.5}$$

If an inverse logit function is used to model a relation of X, $f(X)$, to $\alpha(X)$, then the probability of being protected is modeled

$$\alpha(X) = \frac{1}{1 + \exp(-f(X))}. \tag{15.6}$$

The model $f(X)$ as presented in Dunning (2006) is a two-parameter model $a + bx$. For small assay values, $\alpha(x) \to 0$, and as x gets large, $\alpha(x) \to 1$. Although Model (15.6) looks similar to model (15.10), the interpretation is very different. Model (15.10) is an expression for the probability of developing disease at certain assay and other covariate values, but model (15.6) is an expression for the probability of being protected at a certain assay value.

Combining (15.5) and (15.6) gives a model for the probability that an individual with assay value X develops disease:

$$\Pr(Y_i = 1|X_i = x) = \frac{\omega}{1 + \exp(f(X))}. \tag{15.7}$$

The parameters ω, a, and b can be estimated by standard likelihood methods. Dunning (2006) used a Newton-Raphson algorithm to fit the model.

Given estimates of $\hat{a}$ and $\hat{b}$, suppose that in a trial of a new vaccine candidate in a similar setting, the immunological assays are performed but no clinical outcomes were measured. Let the vaccinated group be denoted by V and the control group by C. Let ω' be the unknown probability of developing disease in the susceptible individuals in the trial. From (15.7), the number of individuals expected to develop disease in the vaccinated group is

$$\sum_{i \in V} \Pr(Y_i = 1) = \sum_{i \in V} \frac{\omega'}{1 + \exp(\hat{a} + \hat{b}x_i)}. \tag{15.8}$$

A similar computation would yield the expected number of cases in the unvaccinated group. In the computation of vaccine efficacy, the value of ω' would cancel in the ratio of expected number of vaccinated and unvaccinated cases. The efficacy of the new vaccine formulation would be predicted by (Dunning 2006)

$$\mathrm{VE}_{new} = 1 - \frac{1/n_v \sum_{i \in V} 1/(1 + \exp(\hat{a} + \hat{b}x_i))}{1/n_c \sum_{i \in C} 1/(1 + \exp(\hat{a} + \hat{b}x_i))}. \tag{15.9}$$

This model assumes that the protective effect at a given titer is the same in the vaccinated and the unvaccinated group.

Forrest et al (2008) used this model to analyze a randomized efficacy study of live attenuated influenza vaccine in young children in the Phillipines and Thailand. They had both an assay for cell-mediated immunity as measured by an IFN-γ ELISPOT and antibodies as measured by HAI.

15.4.3 Household exposure as natural challenge

Storsaeter et al (1998) analyzed a household study nested in placebo-controlled vaccine efficacy trial of acellular pertussis vaccines, a whole cell vaccine all combined with diphtheria-tetanus toxiod compared with diptheria toxin alone. The objectives of the study were

1. to estimate absolute efficacy after household exposure to *B. pertussis* for children with three doses of vaccine compared to placebo recipients;
2. to evaluate possible serological correlates of protection by relating the clinical outcome after household exposure to the antibody levels against PT, PRN, FHA, and FIM;
3. to explore the possible use of post-vaccination anti-pertussis antibody levels as surrogate markers to predict protective efficacy of the whole cell or multicomponent pertussis vaccines.

One of the problems in evaluating correlates of protection is that possibly many of the participants in the study are not even exposed to infection. Examining children with household exposure to pertussis was proposed as a natural challenge experiment.

The outcome Y is 1 if diseased and 0 if not diseased. Let X represent the values of the immunogical assays and possibly vaccination status and other covariates and $g(X)$ is a function of X, for example a linear combination of X, and unknown parameters be estimated. The probability of disease is expressed as a function of X in the logistic model as

$$\Pr(Y = 1|X) = \frac{1}{1 + \exp(-g(X))} \tag{15.10}$$

Of the 329 exposed study participants, 36 had fewer than 3 trial doses. The remaining 293 children were used in computing vaccine efficacy. Of those 209 children fulfilled the generals rules for a valid blood sample for being included in the primary analysis. The guidelines were (1) a pre-exposure ample taken within 4 months of exposure given that it was taken at least 6 months after the third trial dose, or (2) an acute blood sample was accepted if there were no antibody titer rises against either PT, FHA, PRN, or FIM compared to earlier samples. An acute sample was chosen in 125 of the 209 children.

In the nested household study, the efficacy of the five-valent DTaP5 against typical WHO pertussis was estimated at 75.4% (95% CI 59.1 to 85.2) and against any pertussis at 61.8% (95% CI 47.4 to 72.2). In the main trial, vaccine efficacy was estimated at 85.6% and at 77.9%. Fine et al (1988) suggest

Table 15.2. Pertussis cases and vaccine efficacy after household exposure to culture verified *B. pertussis* infection. Only the DTaP5 and DT groups are shown here (from Storsaeter et al 1998).

Clinical definition	Exposed in DTaP5 group $N = 86$ Cases (cult pos)	Exposed in DT group $N = 74$ Case (cult pos)	Vaccine efficacy (95% CI)
Cough 1 day or more and positive lab criteria	28 (13)	63 (43)	61.8 (47.4 –72.2)
Cough 21 days or more and positive lab criteria	21 (11)	60 (43)	69.9 (55.6 – 79.6)
Spasmodic cough 21 days or more (WHO)	14 (10)	49 (36)	75.4 (59.2 – 85.2)

that the more intense and longer exposure in households could result in the commonly observed lower efficacy of pertussis vaccines measured in household based studies.

Storsaeter et al (1998) analyze the data using the arbitrary units obtained in the IgG ELISAs. They also dichotomized the IgG ELISA units in 'Low' (0 to <5 units) and 'High' (≥5 units). The results in the paper focus on the dichotomized analysis. The logistic regression model using the WHO definition and the dichotomized titers was

$$g(x) = 0.675 - 1.12\text{PT} - 1.992\text{FIM} - 1.589\text{PRN} + 1.993(\text{PT} \times \text{FIM}) \quad (15.11)$$

The vaccine group of the child and anti-FHA titer were not statistically significant and not included in the final models. That the vaccine group was not statistically significant suggests that the immunological measures in the model might be considered as fulfilling the Prentice criteria for a surrogate. Based on the WHO definition, the model predicts an attack rate in those with all three values Low as 66.3%. For those with all three values high, the model predicts an attack rate of 11.0%.

Using the method for household exposure to infection as a natural challenge is not feasible in meningococcal vaccine studies because of the low secondary attack rate (Andrews et al 2003).

15.5 Framework for confidence in a biomarker

15.5.1 Correlates of risk

In a series of papers, Qin et al (2007), Gilbert et al (2007), Gilbert and Hudgens (2008a), and Qin et al (2008), propose a framework for assessing immunological correlates of protection in vaccine trials. The framework is based on

the methods of Prentice (1989) and Frangakis and Rubin (2002). The framework delineates different levels of confidence in immunological markers. They in particular distinguish correlates of risk and surrogates of protection. The primary outcome of interest could be clinical disease, infection, or a postinfection outcome. For the discussion here we use VE_S to denote the clinical vaccine efficacy measure of interest and assume it is based on a binary clinical outcome, either infection or disease.

The first, and lowest, level of confidence is a correlate of risk. An immunological measurement that predicts a clinical end point in a particular population is a correlate of risk (CoR). Many vaccine studies have shown that antibody titers correlate with risk of infection or disease. Children with higher immune response to varicella vaccine had lower incidence of chickenpox disease (White et al 1991). Participants with higher immune response to hepatitis B vaccine had a lower incidence of hepatitis (REF). To validate an immunological measurement as a correlate of risk, an association must be observed between these measurements and the clinical end point. Various statistical approaches such as fitting regression models can be used to fit the data for the clinical end point of interest to the immunological measurement (Storsaeter et al 1998, Chan et al 2002, Dunning 2006). The immunological measurement must have a source of variability to be used in the regression models.

If the individuals in the study population have no previous exposure to the infection, they would generally have zero or near zero immune measurements for the infectious agent of interest. Then the correlate of risk can only be evaluated in the vaccinated people. In some diseases in which repeated exposure occurs with the development of partial immunity, such as malaria, or repeated exposure with similar strains, such as influenza, an immunological measurement could be positive and have variability in the unvaccinated people as well as the vaccinated people. For such infectious diseases, the correlation of risk can be evaluated in both the unvaccinated people and the vaccinated people. That is, the clinical outcome of interest can be regressed on the immunological measurements in both the unvaccinated and the vaccinated groups. However, in most vaccine studies, the correlation between immune measure and outcome can be established only in the vaccinated group.

15.5.2 Surrogates of protection

The next two levels of confidence are called surrogates of protection. A surrogate of protection is a correlate of protection, that is, it correlates with a clinical endpoint in some population. In addition, it must predict the level of protective efficacy of the vaccine based on comparison of the immunological measurements in the vaccinated and unvaccinated groups. It only makes sense to evaluate an immunological correlate as a potential surrogate of protection if in fact the vaccine is shown to have a protective effect, that is $VE_S > 0$. Qin et al (2007) differentiate surrogates of protection that predict vaccine efficacy

Table 15.3. Definitions of three levels of an immunological correlate of protection (Gilbert, Qin, and Self 2007)

Term	Definition	Framework for assessment	Analytic method
CoR (Correlate of risk)	An immunological measurement S that correlates with the study endpoint Y measuring vaccine efficacy in a defined population	Vaccine trial (efficacy or proof of concept) or epidemiological study	Regression models
Specific SoP (Surrogate of protection for the same setting)	An immunological measurement that is a CoR within a defined population of vaccine recipients and satisfies either:		
SoP^S (Statistical surrogate of protection for the same setting)	Relation between immunological measurement S and endpoint Y is the same in the vaccine and placebo groups	Single large efficacy trial	Statistical surrogate framework
SoP^P (Principal surrogate of protection for the same setting)	The immune response S satisfies average causal necessity and average causal sufficiency as described in Section 15.5	Single large efficacy trial	Principal surrogate framework
General SoP (Surrogate of protection for new setting)	An immunologic measurement predictive of vaccine efficacy in different settings, such as human populations, viral populations, vaccine lots)	Multiple trials and/or post-licensure studies	Meta-analysis

for the same setting as the source of the data from surrogates of protection predicting efficacy for other settings. The same setting would include a similar population, the same infectious agent, and the same vaccine product. A new setting could be a new population, different strains of the infectious agent, or different vaccine products. They call the former a level 1 surrogate of protection and the latter a level 2 surrogate of protection. Sadoff and Wittes (2007) suggest that the two levels of surrogates of protection be called specific and general surrogates of protection, and we adopt that convention here (Table 15.3).

The specific surrogates of protection are further classified as statistical surrogates of protection (SoP^S) and principal surrogates of protection (SoP^P). The statistical surrogates of protection satisfy the Prentice (1989) criteria for a surrogate described in Section 15.2.1. The data required to evaluate an immunological marker as a statistical surrogate of protection will be available in most clinical vaccine studies if there is considerable variability of the immunological measurement in the control participants. If there is not much variability

in the control group, then it is difficult to evaluate an immunological marker as a statistical surrogate of protection.

The principal surrogates of vaccine protection are based on the principal surrogates proposed by Frangakis and Rubin (2002). The specific principal surrogates of protection are defined by fixed values of the immune response if assigned vaccine, shown more formally in the next section. Let $S(1)$ be the response that an unvaccinated subject would have if vaccinated. Let Y be the 0,1 outcome of being infected or not, Z be the 0,1 assignment to vaccine or control. For a specific principal surrogate of protection, one needs to estimate

$$\mathrm{VE}(s1) = 1 - \frac{\Pr[Y = 1|Z = 1, S(1) = s1]}{\Pr[Y = 1|Z = 0, S(1) = s1]}. \tag{15.12}$$

The definition in expression (15.12) implies that the vaccine efficacy at the immune response level $s1$ is the relative reduction in the risk for groups of vaccinees with immune response $s1$ compared with their risk if they had not been vaccinated. The problem is that in people in the control for whom $Z = 0$, the value of $s1$, the surrogate value under vaccination is not observed.

To asses whether an immunological measurement is a specific principal surrogate of protection, knowledge about $S(1)$ is needed. That is, one needs to be able to predict the immune response that an unvaccinated participant would have had if vaccinated. The two approaches proposed by Follmann (2006) to assess what the immune response would have been in the control participants are possible strategies to evaluate a principal surrogate of protection (Section 15.6.5). An immunological measurement is a specific principal surrogate of protection if two conditions are met. First, groups of vaccinees without responses or with the lowest response levels have a risk equal to that had they not been vaccinated. Second, groups of vaccinees with sufficiently high immune response levels have a risk lower than that had they not been vaccinated.

Although it is useful to understand the relation of immune responses to protection against infection and disease within a particular setting, the goal of identifying surrogates of vaccine protection is to replace large scale phase III trials using clinical outcomes with immunological measurements in new settings and for new vaccines . For example, immunological measures of hemagglutination titer are used to approve the new influenza vaccines each year in Europe. To demonstrate that an immunological marker is a general surrogate of protection requires more stringent data requirements than the specific surrogate of protection. It is actually quite difficult without numerous, likely untestable assumptions. To show that an immunological marker is a general surrogate of protection requires that it predicts vaccine effects on risk across different populations, for different strains, and different vaccine products. One possible approach would be to use meta-analysis combining information from several studies.

15.5.3 Example: pertussis vaccine efficacy revisited

Kohberger et al (2008) evaluated the validity of the Storsaeter et al (1998) model (Section 15.4.3) based on the Gustafson et al (1996) study in terms of the statistical criteria for the validity of surrogate endpoints. They also examined the predictive ability of the model using clinical efficacy data from a different pertussis vaccine efficacy study conducted in Sweden (Olin et al 1997). They call the two studies Sweden I and Sweden II. They examined how five components of acellular pertussis vaccines performed as statistical surrogates of protection against clinical pertussis.

Storsaeter et al 1992

15.6 Evaluating principal surrogate endpoints

15.6.1 Set-up

Gilbert and Hudgens (2008a) define statistical and principal surrogates of protection formally. Their approach is for specific surrogates of protection and evaluates the immunological marker for the same or similar setting as the trial. They introduce an estimand for evaluating a principal surrogate called a causal effect predictiveness (CEP) surface. The causal effect predictiveness surface quantifies how well vaccine effects on the immunological marker predict causal vaccine effects on the clinical endpoint. The CEP surface can be used to compare the surrogate value of several immunological markers.

Consider a randomized, double blind vaccine trial. Assignment is denoted Z, $Z = 1$ for vaccine and $Z = 0$ for control, the discrete or continuous immunological surrogate is S measured at fixed time t_0 after assignment to vaccine or control, and the binary clinical endpoint is Y ($Y = 1$ for disease or infection, 0 otherwise). GIlbert and Hudgens (2008) include an indicator $V = 1$ to denote whether participants are still disease-free at t_0. Later they assume that for any individual, the value of V is the same under vaccine and control. To simplify the presentation, here we will assume that everyone is disease-free at t_0, and drop the notation.

They consider a two-phase outcome dependent case-cohort sampling design (Prentice 1986). A case-cohort study is a case-control study in which the source population is a cohort and every person in the cohort has an equal chance of being included in the study as a control, regardless of how much time that individual has contributed to the person-time experience of the cohort (Rothman et al 2008). In phase one of the study, baseline covariates X are measured on everyone, and in phase two,a baseline covariate(s) W is measured for all or most of the cases, participants with $Y = 1$, and for a random sample of those participants who did not develop disease, $Y = 0$. The candidate immunological surrogate S is measured on everyone for whom W is measured. The indicator δ denotes whether W is measured. Of course, W and

S could be measured on everyone, but it is not necessary in the case-cohort study. For vaccine trials, S and W can be measured after the trial using stored specimens (Nosten et al 1996, Ballou et al 1995)

15.6.2 Defining Surrogates of Protection

Using this notation, a statistical surrogate of protection defined by Frangakis and Rubin (2002) is evaluated by comparing the risk distributions

$$\text{risk}(s|Z=1) \equiv \Pr(Y=1|Z=1,S=s)$$
$$\text{risk}(s|Z=0) \equiv \Pr(Y=1|Z=0,S=s).$$

If for all values of S, $\text{risk}(s|Z=1) = \text{risk}(s|Z=0)$, then the immunological marker S is a statistical surrogate of protection for the clinical endpoint. As described above, the problem with this approach is that what is measured is a mixture of the causal vaccine effects and differences between participants who are infected in the vaccine and unvaccinated groups with values of $S=s$.

A principal surrogate of protection is defined using the notation of potential outcomes in causal inference (see Chapters 1.4 and 9.3.2). Denote the potential clinical endpoint by $Y(Z)$ and the potential value of the immunological marker by $S(Z)$ under vaccine assignment Z. The full potential data are iid copies of $(Z_i, X_i, \delta_i, \delta_i W_i, S_i(1), S_i(0), Y_i(1), Y_i(0))$, $i = 1, \ldots, n$, assuming no drop-out. The usual key assumptions of no interference between units (SUTVA) and independence of treatment assignment from the potential outcomes, e.g. randomization, are made. An immunological marker S is a principal surrogate endpoint if, for all $s_1 = s_0$ the following two risks are equal:

$$\text{risk}_{(1)}(s_1, s_0) \equiv \Pr(Y(1)=1|S(1)=s_1, S(0)=s_0) \tag{15.13}$$
$$\text{risk}_{(0)}(s_1, s_0) \equiv \Pr(Y(0)=1|S(1)=s_1, S(0)=s_0) \tag{15.14}$$

The contrast of the two risks measures a population-level causal vaccine effect on Y for participants with the potential immunological measures $\{S_i(1) = s_1, S_i(0) = s_0\}$.

One requirement for S to be a principal surrogate of protection is in those groups with no causal effect of vaccine on the immunological marker then also the vaccine has no causal effect on the clinical outcome of interest (Frangakis and Rubin 2002) This property Gilbert and Hudgens (2008) call average causal necessity. They further propose that it may be sufficient to show that an immunological marker value above a certain level is sufficient to protect against the clinical outcome of interest, a property called average causal sufficiency. For example, the difference $s_1 - s_0 > C$, where C is some antibody titer or cell-mediate immune response level, could be sufficient to protect against clinical disease, assuming the $s_1 \geq s_0$. Then $\text{risk}_{(1)}(s_1, s_0) < \text{risk}_{(0)}(s_1, s_0)$.

15.6.3 Causal effect predictiveness surface

The causal effect predictiveness (CEP) surface is defined as a contrast, such as the difference, between the two risks in (15.13) and (15.14)

$$CEP^{risk}(s_1, s_0) \equiv \text{risk}_{(0)}(s_1, s_0) - \text{risk}_{(1)}(s_1, s_0), \tag{15.15}$$

where they also consider other contrasts.

The surrogate value of an immunological marker is defined as its capacity reliably to predict the population level causal effect of vaccination on the clinical endpoint. The surrogate value can be quantified by the nearness of the CEP value to 0 for values of the immunological measure under vaccine s_1 close to the immunological measure under control, s_0, and by how the CEP value, that is the difference in the risk under vaccine and control increases as the difference in the two potential immunological measures under vaccine and control, $s_1 - s_0$ increases. Two different immunological markers can have different surrogate values based on differing CEP surfaces.

The marginal CEP curve is defined as a contrast, such as the difference, of the two risks in (15.13) and (15.14) where the risk depends only on the potential immunological marker under vaccine s_1, not also on s_0. When in all participants in the control group, the immunological measure has a 0 or constant value, called the constant biomarker case, then the CEP surface equals the marginal CEP surface.

As defined by Frangakis and Rubin (2002), an associative measure of a principal surrogate of protection is how large the difference is in the potential outcomes under vaccine and control in people whose potential measures of S are *different* under vaccine and control. A dissociative measure of a principal surrogate of protection is how large the difference is in the potential outcomes under vaccine and control in people whose potential measures of S are *same* under vaccine and control. Intuitively, one would want a principal surrogate of protection to have more of an associative measure than a dissociative measure.

Gilbert and Hudgens (2008) suggest functions of the CEP surface that summarize the surrogate values of an immunological marker. The proportion associative effect, PAE^{ω}, is defined by the ratio of the expected associative effect divided by the sum of the expected associative effect and the expected dissociative measure. The question is to what extent the expected associative effect is outweighed by the expected dissociative effect. If PAE^{ω} is in the range [0,0.5], then the immunological measure may have no surrogate value. If PAE^{ω} is in the range (0.5,1.0], then it may have some surrogate value.

15.6.4 Estimating the CEP surface

When the immunological marker has 0 or constant value in the control, it is difficult to evaluate it as a statistical surrogate of protection, as described in Section 15.5.2, though it can be evaluated as a correlate of risk. However,

in this special case, an approach can be taken to estimate the CEP surface and marginal CEP curve. The problem in estimating the CEP surface is that we do not observe the immunological responses S under both vaccine and control. So now include the baseline covariates X and W in the expressions for the two risks in (15.13). Under the assumption of SUTVA and independent assignment mechanism, (15.14):

$$\text{risk}_{(1)}(s_1, s_0, x, w) = \Pr(Y = 1 | Z = 1, S = s_1, S(0) = s_0, X = x, W = w) \tag{15.16}$$

$$\text{risk}_{(0)}(s_1, s_0, x, w) = \Pr(Y = 1 | Z = 0, S(1) = s_1, S = s_0, X = x, W = w) \tag{15.17}$$

We would be able to estimate the two risks if we knew the potential outcomes $S_i(Z)$ of participants if they had been assigned the opposite treatment.

If the response to the immunological marker is 0 or constant in everyone if in the control group, then $\text{risk}_{(1)}(s_1, s_0, x, w)$ can be estimated from the observed data. However, the potential value $S_i(1)$ of the immunological marker if vaccinated in those participants who received control needs to be determined to be able to estimate the CEP surface.

Assume through an augmented vaccine trials design (Follmann 2006) (Section 15.6.5), a baseline covariate W predictive of the immunological measure $S(1)$ is measured in both treatment arms. Then a model predicting $S(1)$ from X and W can be fit in the participants in the vaccine group and used to predict the potential value of the immunological measure $S(1)$ for participants in the control group. Details of estimation and inference as well as a test of whether an immunological measure has any surrogate value are in Gilbert and Hudgens (2008). The surrogate marker value of $S(1)$ for people in the control group is treated as missing data. The likelihood contribution for a person in the control group is obtained by integrating the risk over the conditional cumulative distribution function of $S(1)|X, W$ in the vaccinated group.

15.6.5 Augmented designs to assess immune response

Follmann (2006) proposed two augmented vaccine trial designs to help determine whether a particular immune response to a vaccine is actually the causal factor in reducing the infection rate in the vaccinated compared to the unvaccinated group. The first approach involves vaccinating everyone before baseline with an irrelevant vaccine. For example, in a pneumococcal vaccine trial, one might vaccinated the control group with a meningococcal vaccine. Then randomization ensures that the relationship between the immune responses to the meningococcal and pneumococcal vaccines observed in the vaccine group is the same as that that would have been observed in the control group. The response to the pneumococcal vaccine in individuals in the control group can be inferred from their response to the meningococcal vaccine and a prediction model from the vaccine group.

In the second approach, all uninfected participants in the control group are vaccinated with the pneumococcal vaccine at the end of the trial and their immune responses are recorded. Then one assumes that the immune response that they have at the end of the trial is the response they would have had if vaccinated at the beginning of the trial. By comparing with the full distribution of immune responses in the vaccinated group, because of randomization, one can infer what the distribution of immune response in the infected participants in the control group would have been.

Qin et al (2008) develop details of using case-cohort sampling and a Cox proportional hazards model to assess surrogate endpoint candidates in vaccine trials as developed by Gilbert and Hudgens (2008) using the two different augmented vaccine trial designs suggested by Follmann (2006).

15.7 Further considerations

Role of baseline transmission in evaluating a surrogate. Suppose that exposure to infection and the baseline attack rates are higher in some populations than others. Would this affect the surrogate value of an immunological marker? To what extend is the development of the principal surrogate of protection dependent on an all-or-none model of protection?

Rubin (2004) discusses the use of immunological surrogates in monkey trials of anthrax vaccines and extensions to humans. The causal effect of dose contradicts the Prentice criteria.

15.8 Examples

15.8.1 Pertussis vaccines in Sweden

15.9 Waning of antibodies

Gilks et al (1993) estimated the waning of antibody titers with a random-effects models for longitudinal data using Gibbs sampling. This approach can be used to determine schedules for booster shots if it is known what level of antibodies are protective.

15.10 Carriage as outcome in pneumococcal trials

Nasopharyngeal carriage is being considered as an endpoint for trials of pneumococcal vaccines. The project, called PneumoCarr, is based in Finland. The goal of the PneumoCarr project is to establish reduction of colonization as part of the licensure process. The vaccine efficacy measure is called VE_{col}. A second goal of the project is to identify serological correlates of VE_{col} Information is

available from the website http://www.ktl.fi/roko/pneumocarr/index.html. A new endpoint in pneumococcal vaccines trials is need. Invasive pneumococcal disease is a rare event. It cannot be studied in detail, its diagnosis is dependent on local medical practices and . it requires special equipment and training. Acute otitis media as the primary outcome is problematic. Specific diagnosis of pneumococcal bacteria as the infectious agent causing the otitis media needs special procedures to obtain a bacteriological outcome. Otitis media is perceived differently in different parts of the world. Pneumonia as an outcome is problematic because diagnosis and definition of a case is not specific. A lung aspirate is required, which is simply not feasible, so there is no bacteriologic endpoint.

(REFERENCES) Using nasopharyngeal carriage as an endpoint in vaccine trials makes sense for several reasons. It is the most important endpoint because preventing carriage will prevent all of the other endpoints directly in the person vaccinated and it will prevent transmission to others, since carriage is the source of infection to others. Nasopharyngeal carriage as a pneumococcal endpoint can teach about other mucosal infections. Comparing the relative incidence of the various candidate outcomes, carriage is the most frequent by logs.

Using nasopharyngeal carriage as the outcome seems feasible. It is the most common endpoint and the most accessible endpoint. Nasopharyngeal carriage is abundant both before and after introduction of vaccine. It permits feasible follow-up of dynamics after introduction of vaccines, such as reduction in carriage, and development of antibiotic resistance.

However, so far the known predictors for protection against invasive pneumococcal disease do not predict protection against carriage, spread and mucosal infections. The main caveat with nasopharyngeal carriage as an endpoint is the difference in disease potential of different serotypes to various infection sites.

Three type of predictors can be used to measure the effect of vaccination on nasopharyngeal carriage. Prevention of new acquisitions measures direct protection. Prevalence of carriage is a measure of indirect protection. Density of carriage is a measure of direct and indirect protection.

Modeling of pneumococcal studies based on longitudinal household and school studies is presented in Chapter 11. Modeling results reveal that the variation between studies and between populations is enormous. Modeling can provide at least hypothesis generating findings. The measure VE_{col} is linked to transmission. It is a direct measure of functional immunity and serotype specific. It allows prediction of population level effects, assessment of non PCV pneumococcal vaccines

Problems

15.1. Estimating predicted efficacy
(a) Show that the estimated model (15.11) for the WHO definition of pertussis gives the predicted attack rates 66.3% and 11.0%.
(b) Another case definition in Storsaeter et al (1998) was whether the study child was laboratory positive for pertussis and had at least one day cough during the follow-up period. A non-case was a study child who was laboratory negative or laboratory positive but without cough. For this definition, the fitted model was

$$g(x) = 2.003 - 0.146\text{PT} - 1.548\text{FIM} - 1.990\text{PRN} + 1.148(\text{PT} \times \text{FIM}) \quad (15.18)$$

What are the predicted attack rates in the Low and High groups?

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