Statistical Methods in Medical Research 2004; **13**: 1–26

Endpoints in vaccine trials

Michael G Hudgens, Peter B Gilbert and **Steven G Self** Statistical Center For HIV/AIDS Research and Prevention, Program in Biostatistics, Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA

In this paper we discuss statistical considerations regarding endpoints in preventive vaccine trials. Brief discussion is given to preclinical, Phase I, and Phase II trials, with the bulk of attention paid to endpoint choice and analysis in Phase III efficacy trials. In addition to traditional efficacy measures of vaccine effects for immunized individuals, consideration is given to waning, strain specific efficacy, correlates of protective immunity, post-infection endpoints, and cluster randomized trials.

1 Introduction

Vaccines are widely considered one of the greatest achievements in public health, having had a dramatic impact on the prevalence of several infectious diseases including smallpox, poliomyelitis, and measles. Since the 1940s, clinical trials have become critical for evaluating new vaccines as well as other prevention and treatment strategies in combating human diseases. Today the randomized, controlled trial is the gold standard for providing scientific evidence regarding the efficacy of a candidate vaccine.[1] In general, vaccine clinical trials proceed in an ordered sequence of studies denoted as Phases I, II, and III. Phase I trials typically involve a small number of participants ($n \sim 10$–100) and seek to evaluate vaccine safety and *tolerability* over different dosages or regimens. Preliminary assessment of vaccine immunogenicity (that is, the vaccine's ability to stimulate an immune response) may also be possible in a Phase I trial. Phase II trials are usually larger ($n \sim 100$–500), allowing more accurate characterization of safety and immunogenicity. For vaccine candidates that are safe and immunogenic in Phase I and II trials, Phase III trials ($n \sim 1000$–$100\,000$) are employed to evaluate efficacy of the vaccine within the population of interest. Vaccines that prove to be safe and efficacious in Phase III trials may be licensed by the appropriate regulatory agency. Given licensure, nonrandomized observational studies, sometimes called Phase IV studies, are typically employed to assess vaccine effectiveness and safety in the field.

In this paper we discuss statistical considerations regarding endpoints in vaccine trials. An *endpoint* is generally defined as a measurement determined by a trial objective that is evaluated in each study subject.[2] Like all clinical trials, careful definition of endpoints prior to study initiation aids in trial design and hypothesis formulation, provides guidance in the analysis of the data upon trial completion, and enhances the

Address for correspondence: MG Hudgens, 1100 Fairview Ave N, MW-500, PO Box 19024, Seattle, WA 98109-1024, USA. E-mail: mhudgens@scharp.org

 10.1191/0962280204sm356ra

Techset Composition Ltd, Salisbury Doc: g:/Journals/Hodder/smmr/SM356RA.3d Manu No. 0000
Printed: 13/1/04 Page#: 26 page(s) Opp: Task: 3B2 Version: 7.51f/W (Mar 4 2002)/TechRef:9.01/H (March 20 1997)

credibility of the results.[3,4] Considerations regarding the selection of endpoints should include 1) the rate of occurrence of the endpoints in the population under consideration, 2) the relevance of the vaccine effect on the endpoint, and 3) reliability in measuring the endpoint.[5] Thus the available sample size is critical to endpoint choice. As such, the remainder of the paper is organized by trial phase. Emphasis is given solely to endpoints in preventive vaccine trials, recognizing that therapeutical vaccines can require a different set of considerations.[6,7] Some attention is also given to estimands, estimators, and analysis methods corresponding to the endpoints under discussion.

2 Preclinical studies

The earliest phases of vaccine development begin with investigation of candidate vaccines in animals (*in vivo*) and in laboratories (*in vitro*).[8] Animal studies are designed to assess safety, immunogenicity and, provided an appropriate challenge model exists, efficacy.[9] In these studies, the assessments are typically viewed as tests of biological concepts with attention paid more to qualitative than quantitative outcomes. For example, preliminary tests of the concept of a live attenuated HIV vaccine in macaques have clearly demonstrated the reality of reversion of attenuated virus to pathogenic strains, although few would attempt to infer specific kinetics of reversion in humans from those in macaques. With respect to immunogenicity outcomes, animal models played a key role in determining the existence of three serotypes of polio virus.

In some circumstances, animal models may be used to compare or rank different vaccine regimens by their immunogenicity or efficacy profiles and to select the higher ranking regimens for clinical evaluation in humans. This more demanding use of animal models is analogous in some respects to the statistical problem of surrogate endpoints in that rankings in animals are used as surrogates of (or predictors for) corresponding rankings in humans. This use of animal models may be considered when the number of vaccine regimens to be screened is so large that evaluation in humans is impractical. This can occur when screening large numbers of candidate antigens derived from different regions of the pathogen's genome[10] or when considering vaccine regimens formed from combinations of a number of component antigens. Challenges of manufacturing the multiplicity of reagents that are suitable for human testing may also motivate this use of animal models for ranking and selection. Just as in the classical surrogate endpoint problem, there is uncertainty as to how well the ranking of vaccine candidates in animal models is preserved in humans. However, unlike the classical problem, there is greater potential to assess accuracy of predictions in this setting. Multiple candidate regimens can be selected for evaluation in human trials providing a direct assessment of how well rankings in animals correspond to those in humans within the 'population of vaccine regimens' considered.

Preclinical vaccine development also entails *in vitro* studies, such as assessment of quality control of the manufacturing process (for example, lot-to-lot variability, stability and sterility testing) and validation of immunogencity assays to be used in subsequent clinical trials. *In vitro* studies are also a fundamental component of vaccine design. For example, neutralization assays are employed in the serotypical characterization of new circulating strains of influenza, which in turn affects vaccine formulation.

Limitations of the animal model, such as differences in susceptibility to the pathogen of interest and in the major histocompatibility complex (MHC) compared to humans, can render *in vitro* studies critical to vaccine design. In this regard, epitope-driven vaccine development[11] relies on sensitive and specific *in vitro* T-cell assays to confirm predicted epitopes from bioinformatics mining of pathogen genomes for vaccine components.

3 Phase I trials

Phase I vaccine studies can range from small trials involving the first use of a vaccine candidate in humans to larger trials ranging over different doses, immunogens, constructs, or administrative schedules. Assessment of safety is often the primary objective, such that participants are usually healthy adults at low risk of acquiring the infection or disease of interest. Depending on the setting, enrolment may be limited to include only 'naive' volunteers (as determined by serology), or, conversely, only individuals having prior infection with the pathogen of interest. Vaccine-induced immunogenicity typically constitutes a secondary objective. These trials are often open-label and nonrandomized;[8,12] moreover, strict double-blinding may not be feasible due to operational constraints such as staggered enrolment in dose ranging studies or different delivery mechanisms.

Safety endpoints are typically classified into reactogenicity and vaccine-related adverse experiences (AEs), of which the most severe are reported as serious adverse experiences (SAEs). Reactogenicity is often further divided into systemic (for example, fever, malaise, myalgia) and local (for example, pain, tenderness, induration at the injection site) symptoms. Given the number of potential safety endpoints, a multiplicity adjustment might be considered[13] for confidence intervals or tests of differences in rates from a comparable control group[14,15] (for example, a placebo arm if present). On the other hand, since the outcomes pertain to safety, in the interest of sensitivity one may want to proceed conservatively by avoiding a multiple comparisons adjustment and accepting the possibility of an inflated false positive rate.[4,8] In this case, good clinical and statistical judgement should be employed in weighing the unadjusted p-values such that an innocuous vaccine is not incorrectly deemed unsafe. Mehrotra and Heyse[14] suggest a two-step false discovery rate (FDR) approach that offers a balance between no adjustment and 'over' adjustment, and illustrate the methodology with application to adverse event data from three vaccine trials.

Owing to the small sample sizes of early vaccine trials, the novel statistical issues regarding immunogenicity may pertain more to rapidly evolving measurement technology rather than to trial design. For example, in the development of an HIV vaccine, induction of a strong and durable CD8+ CTL response to HIV-1 has become an important immunogenicity outcome. Historically, CTL activity has been measured by a ^{51}Cr chromium release cytotoxicity assay;[16–20] more recently the ELISpot assay has begun to play a critical role in the assessment of immunogenicity.[21,22] For either assay, methods for assessing whether a participant has a positive qualitative response are usually based on reasonable but *ad hoc* approaches. Recently Self *et al.*[23] considered statistical methods that are helpful in the analysis of ELISpot data as well as the design of the assay. Improvement over *ad hoc* positive criteria using statistical methods has

also been proposed[24] for analysis of a proliferation assay in the context of developing a vaccine for herpes simplex virus type 2. In general, the employed statistical methodology in this setting strives to minimize false positives and negatives while facilitating investigator control of these two error rates. In addition to functioning as vaccine trial endpoints, assays measuring immune response are critical for mechanistic studies and as potential correlates of protection (see section 5.6 below).

4 Phase II trials

The primary objectives of Phase II trials entail further characterization of safety and immunogenicity,[8,25,26] usually in order to attain the broader goal of identifying the most promising preparation, dose, and schedule to be tested in Phase III trials. Typically randomized, double-blind, and placebo-controlled Phase II trials enrol individuals from the target population for whom the vaccine candidate was developed.[8,26,27] By virtue of these differences and an increase in sample size, Phase II trials allow more precise characterization of safety and immunogenicity than do Phase I trials. These trials are usually powered to test for or rule out putative or known clinically meaningful differences in the distributions of immunogenicity and safety endpoints[26] such that information on background rates and variability of endpoints is needed for sample size calculations. Note that evaluation of safety in Phase I and II trials is limited to detection of relatively common endpoints; rare adverse experiences, such as febrile convulsions following measles vaccination, often can only be detected in large, post-licensure Phase IV observational studies.[27]

Immunogenicity endpoints require careful consideration of several factors. Determination must be made as to which types of immune responses (for example, humoral, cellular, mucosal, and so on) are of interest and at which time points they are measured. Antibody response to the target pathogen usually constitutes the primary immunogenicity endpoint of Phase I and II vaccine trials. A quantitative assay such as the enzyme-linked immunosorbent assay is employed to measure antibody response, with comparisons between unvaccinated and vaccinated groups entailing contrasts in proportions responding (using a pre-determined threshold believed to correlate with protection[28]) or geometric means (often assuming log-normality). Regression methods can be employed to examine relevant covariates as potential effect modifiers of the antibody response to vaccine (for example, age or pre-existing immunity) and to control for confounding. For example, Moulton *et al.*[29–31] suggest using percentile regression techniques or a mixture gamma model approach that allows for left censored observations below the assay's limit of detection as well as a subpopulation of nonresponders.

While historically immunogenicity endpoints have been defined solely in terms of antibody, presently it is possible to measure additional immune responses (for example, mucosal or cellular) such that, like safety, the issue of multiplicity arises. At a minimum, one might argue that multiple measurements of a specific type of immune response constitute a family of hypotheses that warrants a multiplicity adjustment. For example, cellular immune responses are often measured by the ELISpot assay where cytokine release to several different peptide pools of the antigen of interest (say HIV) is evaluated; here an adjustment should be made for the multiple peptide pools

considered.[21] On the other hand, endpoints reflecting the different components of the immune system might be considered separately without a multiplicity adjustment (for example, prime/boost HIV vaccine regimens that attempt to induce both cellular and humoral responses). Depending on the number of genotypes or serotypes of the pathogen, similar considerations apply to the analysis of immune responses measured to several strains of the pathogen. The purpose of an analysis and the number of endpoints affect the choice of multiple comparisons adjustment procedure. Methods that control the familywise error rate (FWER) may be preferable when the goal is to identify any vaccine effect (for example, the ELISpot example cited above), whereas methods that control the FDR may be more suitable when the goal is to identify the set of endpoints on which there is a vaccine effect (for example, to characterize the specific pathogen strains to which the vaccine responds). When the number of endpoints is large, FWER-controlling methods are usually highly conservative; in this case FDR-controlling methods are more powerful and control Type I and II errors in a more balanced manner.

Typically, immune response endpoints are defined at some fixed time point after the last vaccination. An alternative is to define an endpoint that summarizes several longitudinal immunogenicity responses. For example, one of the primary endpoints in AVEG 202/HIVNET 014,[16] a Phase II trial of a canarypox vector based HIV vaccine candidate, was defined as having at least one CD8+ CTL response at either day 98 or 182 post-randomization. In the presence of ignorable missingness, maximum likelihood methods can be employed to analyze such an endpoint.[32]

Phase II trials may also provide partial information on vaccine efficacy, in which case the moniker 'Phase IIb' may be used.[25] In settings where an infectious disease occurs with sufficiently high incidence (for example, rotavirus[33,34]), Phase IIb trials can provide a preliminary assessment of efficacy. In this case, endpoint considerations are similar to those of a Phase III trial (section 5). Estimates of vaccine efficacy may also be gleaned from challenge studies wherein volunteers are deliberately inoculated with the target pathogen; such an approach has been employed in testing vaccines for cholera, malaria, influenza, and typhoid fever.[25] Because exposure is under the control of investigators, these trials can use classical experimental designs[35] to assess directly many vaccine effects of interest that cannot be observed in typical Phase III efficacy trials. In particular, using infection or disease as the endpoint, challenge studies can provide estimates of strain-specific efficacy. For example, Levine *et al.*[36] describe a challenge study designed to assess the efficacy of recombinant live oral cholera vaccines to different biotypes and serotypes using diarrhoea as the endpoint. Additional endpoints reported include shedding of the challenge and vaccine strains, which could be considered surrogate markers of secondary transmission of the virus and vaccine, respectively. Immune responses can also be measured closely prior to infection in these trials, providing important information regarding the establishment of an immune correlate of protection (see section 5.6 for further discussion).

Phase II trials may also provide sufficient evidence of efficacy if a correlate has previously been established as a valid surrogate endpoint for infection or disease. This approach might be employed when introducing an efficacious vaccine into a new population[37] or when combining two or more existing vaccines.[38] Even if such a correlate exists and provides adequate evidence to substantiate efficacy, a large Phase III

safety trial may still be necessary since the usual extensive clinical safety data from a well-controlled Phase III efficacy trial will not exist.[38]

5 Phase III trials

Typically the primary objective of a Phase III trial is to estimate efficacy of a candidate vaccine in the population of interest. Like Phase II trials, efficacy trials are usually randomized, double-blind, and placebo-controlled. The ability to randomize participants to a placebo group is dictated by ethical considerations; in particular, whether or not an efficacious vaccine already exists. Unless otherwise specified, we assume a two-arm trial with one arm randomized to the vaccine of interest and the other arm randomized to placebo (typically a vaccine for another disease).

Continuing to assess safety and tolerability of a vaccine candidate is also an objective in Phase III trials. Given the intensive evaluation of reactogenicities and adverse events in earlier trials, safety monitoring tends to be more passive in Phase III trials. Inference concerning safety can be challenging even in large efficacy trials since the goal is to demonstrate a lack of association between safety outcomes and the candidate vaccine.[8] Even if an association is detected, the potential public and individual health benefits of the vaccine require a risk–benefit assessment. As a result, an equivalency or noninferiority approach might be employed wherein one tests the hypothesis that no more than a specified difference in safety profiles exists between placebo and vaccine arms.[26] For rare events, sufficient power for noninferiority will not be feasible, such that subsequent large, simple safety trials or Phase IV observational studies (see section 7) will be necessary. With regard to the former, Horne *et al.*[8] recommend that common, less serious adverse events be monitored in only a subset of participants, while the incidence of SAEs should be closely monitored in all individuals.

The remainder of this section pertains to the evaluation of efficacy in Phase III trials.

5.1 Disease as the primary endpoint

The definition of trial endpoints with respect to efficacy depends on characteristics of the disease and the candidate vaccine. In general, the goal of vaccination is to prevent or ameliorate disease, and *not* necessarily to prevent infection. For example, vaccines for rubella, mumps, measles, and polio have been shown to prevent disease, but not infection.[9] Therefore, the primary efficacy endpoint of Phase III vaccine trials is usually defined with respect to clinically significant disease morbidity or mortality. While the wide range of clinical outcomes from infection may necessitate assessing vaccine efficacy on several endpoints (potentially requiring multiplicity adjustment), we will assume for now that there exists a sole endpoint that measures clinically significant disease.

Vaccine efficacy typically has the form $\mathrm{VE} = 1 - \mathrm{RR}$ where RR denotes the relative risk of disease in vaccinees compared to placebo recipients, that is, $\mathrm{RR} = \mathrm{R_V}/\mathrm{R_P}$ where $\mathrm{R_V}$ and $\mathrm{R_P}$ denote the risk in the vaccine and placebo arms respectively. Given that a risk ratio must be non-negative, it follows that $\mathrm{VE} \in (-\infty, 1]$ with a value of 1 indicating complete protection, 0 representing no effect, and a negative value conveying an increase in risk due to vaccination. Vaccine efficacy is usually defined in terms of

relative attack rates or hazard functions.[37,39] The cumulative incidence or attack rate estimand is defined by

$$\mathrm{VE}_{\mathrm{CI}} = 1 - \frac{\mathrm{CI}_{\mathrm{V}}}{\mathrm{CI}_{\mathrm{P}}} \tag{1}$$

where CI_{V} (CI_{P}) is the cumulative incidence or probability of disease over the course of a trial of duration t in the vaccine (placebo) arm. This measure of vaccine efficacy ($\times 100\%$) indicates the percent reduction in the risk of developing disease during the trial attributable to vaccination. Measuring efficacy by $\mathrm{VE}_{\mathrm{CI}}$ is appropriate if it is believed that the vaccine has an 'all-or-nothing' mode of action whereby the effect of the vaccine is to render some proportion of those vaccinated completely immune while offering the remainder no protection.[39] Alternatively, vaccine efficacy can be defined in terms of the hazard or incidence ratio,

$$\mathrm{VE}_{\lambda} = 1 - \frac{\lambda_{\mathrm{V}}}{\lambda_{\mathrm{P}}} \tag{2}$$

where λ_{V} and λ_{P} are the incidence of disease in the vaccine and placebo groups, respectively; time-dependent generalizations of VE_{λ} are discussed in section 5.8. This measure of vaccine efficacy is appropriate if it is believed the vaccine is 'leaky,' that is, vaccination reduces the hazard of disease by a constant, multiplicative factor that is equal for all vaccinees.[37,39] For a time-constant incidence rate, the two vaccine efficacy measures $\mathrm{VE}_{\mathrm{CI}}$ and VE_{λ} are related by the equation $\mathrm{CI} = 1 - \exp(-\lambda t)$, and are approximately equal for small values of λt.[37] It follows that for rare diseases, use of either estimand $\mathrm{VE}_{\mathrm{CI}}$ and VE_{λ} is approximately correct for both all-or-nothing and leaky vaccines, a useful fact since the vaccine mechanism is frequently unknown.[37] Note that both $\mathrm{VE}_{\mathrm{CI}}$ and VE_{λ} are relative risk measures and, as such, will not necessarily capture all of the information pertaining to the effect of the vaccine. Moreover, absolute differences in attack or hazard rates should also be considered and, in some settings, may provide more practical information from a public health policy perspective.[40]

Regarding estimation of vaccine efficacy, rates of disease in the vaccine and placebo arms can be used to estimate $\mathrm{VE}_{\mathrm{CI}}$ in the absence of censoring. Specifically, let n_{V} (n_{P}) be the number of disease cases in the vaccine (placebo) arm and N_{V} (N_{P}) be the number of volunteers randomized to vaccine (placebo), such that

$$\widehat{\mathrm{VE}}_{\mathrm{CI}} = 1 - \frac{\hat{p}_{\mathrm{V}}}{\hat{p}_{\mathrm{P}}} \tag{3}$$

where $\hat{p}_{\mathrm{V}} = n_{\mathrm{V}}/N_{\mathrm{V}}$ and $\hat{p}_{\mathrm{P}} = n_{\mathrm{P}}/N_{\mathrm{P}}$ are the attack rates. Given an all-or-nothing mechanism and no censoring, the estimator $\widehat{\mathrm{VE}}_{\mathrm{CI}}$ can be viewed as a consistent but biased maximum likelihood estimator (MLE) of $\mathrm{VE}_{\mathrm{CI}}$ since it is nonlinear in the MLEs $\hat{p}_{\mathrm{V}}$ and $\hat{p}_{\mathrm{P}}$.[41] Several bias corrected estimators have been proposed, but in general the bias of $\widehat{\mathrm{VE}}_{\mathrm{CI}}(T)$ is of concern only for smaller (for example, Phase IIb) trials.[41] In the presence of right or interval censoring, lifetable (for example, see Szmuness *et al.*[42]) or

nonparametric maximum likelihood[43,44] estimators of CI_V and CI_P at time t can be substituted into Equation (1) to obtain an estimator of VE_{CI}. While likely not an important factor in the analysis of trials with relatively short follow-up, failure to properly account for censoring in vaccine trials requiring longer follow-up (for example, for diseases such as hepatitis B, HIV, cholera) can lead to biased estimators of VE_{CI}. For VE_λ, empirical estimators are typically employed by using the number of disease cases per person-time follow-up. Under a proportional hazards assumption, VE_λ can be estimated by

$$\widehat{VE}_\lambda = 1 - e^{\hat{\beta}} \tag{4}$$

where $\hat{\beta}$ is the partial likelihood estimate of the log hazard ratio.[45] Standard methods for dichotomous or survival outcomes can be employed for confidence interval estimation, testing, and covariate adjustment of VE_{CI} and VE_λ. For example, Szmuness *et al.*[42] use the logrank test to detect differences in endpoint rates from a hepatitis B efficacy trial. Similarly, in a malaria vaccine efficacy trial, Alonso *et al.*[46] use Cox regression models to adjust the estimated efficacy against infection by age and distance between home and dispensary.

5.2 Infection as the primary endpoint

For diseases with long incubation times such as HIV and tuberculosis, the traditional endpoint of clinical disease morbidity or mortality may not be feasible due to the required duration of follow-up. Thus the more proximal endpoint of infection may be designated primary.[47,48] In this case, VE_λ and VE_{CI} are defined in terms of infection instead of disease with analysis proceeding accordingly. Additionally, a third measure of VE can be defined as

$$VE_\rho = 1 - \frac{\rho_V}{\rho_P} \tag{5}$$

where ρ is the probability of infection given a specified exposure to infection or inoculum.[49] VE_ρ is often referred to as *biological efficacy*.[48,50] By assuming a particular underlying epidemic model, Haber *et al.*[49] propose estimating VE_ρ by Equation (3) if the vaccine modality is all-or-nothing, and by

$$\widehat{VE}_\rho = 1 - \frac{\ln(1 - \hat{p}_V)}{\ln(1 - \hat{p}_P)} \tag{6}$$

if the vaccine is believed to be leaky. Bias corrected variations of Equation (6) given by Chick *et al.*[41] should be considered for smaller trials.

5.3 Post-infection endpoints

In settings where infection is a primary endpoint, also assessing vaccine efficacy against disease is critical. Considering vaccine effects on infection only allows for the possibility of failing to identify vaccines that protect against or enhance disease. Note,

however, that for vaccines against chronic infections assessing effects on disease may require long-term follow-up such that a more immediate analysis might consider vaccine effects on a surrogate endpoint for onset of disease.[50–52] For example, in the first two HIV vaccine efficacy trials, the primary endpoint is HIV infection, while the secondary endpoint focuses on the extent and duration of viremia in infected participants,[53] a putative surrogate marker of progression to AIDS. Issues of validating a surrogate endpoint aside (see section 5.6), analysis of vaccine effects on disease in the context of infection as the primary endpoint presents several statistical challenges.

First, if effective treatments are available for the disease of interest, any vaccine effect on disease progression (or a surrogate marker thereof) may be confounded by providing such treatments to infected trial participants. Fine[54] makes this point in the context of evaluating efficacy of acellular pertussis vaccines. Another example occurs when assessing rotavirus vaccine efficacy to prevent severe childhood diarrhoea in trials where effective oral rehydration therapy is commonly administered at the first symptom of diarrhoea (Jorge Flores, personal communication). Continuing with the HIV example, infected participants may begin antiretroviral therapy (ART), which is known to lower viral load. The problem of confounding can be alleviated by considering a composite endpoint of time until virologic failure above some pre-set threshold or initiation of ART,[55] or by focusing on pretreatment viral load only and employing methods designed to correct for potentially dependent censoring induced by ART. For example, the semiparametric approaches of Rotnitzky *et al.*[56] might be employed wherein the probability of censoring (that is, ART initiation) is modeled as a function of viral load and other factors using logistic regression.

Secondly, one must decide whether the analysis of disease should include all participants or only those who are infected. The former approach enjoys the statistical validity associated with an intent-to-treat (ITT) analysis; see section 5.4 below. Chang *et al.*[57] explore this approach by assigning disease severity scores (for example, see Flores *et al.*[33]) to each incident case and then considering differences in sums of scores for vaccine and placebo arms as a burden-of-illness efficacy measure. Specifically, let $S_{p1}, \ldots, S_{pn_P}$ and $S_{v1}, \ldots, S_{vn_V}$ be the severity scores for the n_P and n_V infected individuals in the trial. Then Chang *et al.*[57] consider

$$T = \frac{\sum_{i=1}^{n_P} S_{pi}}{N_P} - \frac{\sum_{i=1}^{n_V} S_{vi}}{N_V} \tag{7}$$

as an estimator of the net reduction in morbidity per randomized subject. In a similar fashion, one could consider a more traditional, relative risk-based estimator such as

$$\widehat{\mathrm{VE}}_{\mathrm{severity}} = 1 - \frac{\sum_{i=1}^{n_V} S_{vi}/N_P}{\sum_{i=1}^{n_P} S_{pi}/N_V} \tag{8}$$

that is, the percent reduction in morbidity score. Note that by choosing severity scores all equal to one, Equation (8) is equivalent to the attack rate estimator (3). That Equations (7) and (8) have an ITT interpretation follows by noting that uninfected individuals are effectively being assigned a score of zero. This approach is also

appealing in that the estimator provides an overall measure of the net benefit of a vaccine on incidence and disease that avoids issues of multiplicity and selection bias (as discussed further in the following paragraph). On the other hand, potential drawbacks of this approach include not clearly differentiating vaccine effects on infection and pathogenesis as well as the challenge of choosing the severity scores in a meaningful fashion.

As an alternative to an ITT-based approach, one might consider an analysis that contrasts disease severity or progression rates in infected vaccinees and infected placebo recipients only. For example, Vesikari *et al.*[58] present results from a rhesus rotavirus vaccine trial where vaccine efficacy was based on the endpoint of rotavirus diarrhoea. Additionally, the effect of the vaccine on the clinical course of infection was considered by comparing severity (mild, moderate, or severe) between vaccinees and placebo-treated individuals with confirmed rotavirus diarrhoea using Fisher's exact test. Such a comparison using infected participants only should be interpreted with caution since contrasts are being made between two subsets of the original randomization groups whose membership has been determined by a post-randomization event, namely infection, and thus are subject to selection bias. Procedures to test for vaccine effects on disease progression in infected individuals that account for selection bias have been proposed recently in the context of viral load analysis in HIV vaccine trials.[51,52] In particular, selection models can be formulated using a causal inference framework that allows testing for a causal effect of the vaccine on viral load in the basic principal stratum[59] of individuals who would have been infected regardless of randomization assignment. Employing such selection models is especially important in testing for vaccine harm in infected individuals, a phenomenon for which there is precedence and concern.[60–63] Net comparisons of infected vaccinees and placebo recipients that do not account for selection bias could erroneously suggest vaccine-enhanced pathogenesis when in actuality the vaccine is simply protecting individuals with stronger immune systems from infection, against infection with relatively innocuous viral strains, or by some other selective mechanism. Q1

In addition to dealing with selection bias, an analysis of vaccine effects on infection and disease must consider whether these endpoints are co-primary or designated as primary and secondary. A co-primary scenario might entail a joint analysis of both effects while a primary–secondary approach might consider effects on infection and disease separately with due attention paid to the overall significance level α of the trial.[64]

5.4 Intent-to-treat versus per-protocol

The intent-to-treat (ITT) principle generally refers to analysing all randomized participants according to randomization assignment regardless of treatment received or compliance, with follow-up measured from the time of randomization.[7] In general, ITT has become the gold standard in clinical trials since it ensures the validity of testing the null hypothesis of no treatment effect and helps minimize bias such that differences in outcomes between the groups can be attributed to the treatment under study. While therapeutical vaccine trials have typically employed an ITT analysis per FDA recommendation,[7] the norm for preventive vaccine trials has been to take a per-protocol or as-treated approach wherein only fully compliant volunteers with respect to

immunization are included in the analysis of efficacy. Additionally, the endpoint is often defined as disease (or infection) only after some suitable time lag beyond completion of the immunization series to allow for optimal immunity. Such per-protocol analyses have been advocated on the basis of providing information on the intrinsic[40] efficacy of the vaccine after completion of the prescribed regimen. However, like the post-infection endpoint analysis described above, a per-protocol analysis entails comparison of subgroups selected post-randomization, and thus is subject to bias. On the other hand, the ITT approach has been advocated in general[65] as well as within the context of preventive vaccine trials[7,40] for rendering results that are more readily applicable to the population of interest.

Despite differences in these two approaches, Horne *et al.*[7] generally found little difference in ITT and per-protocol vaccine efficacy estimates from several trials reported in the last 20 years. This concordance was attributed to excellent compliance and few endpoints occurring during the period of immunization typical of the preventive vaccine trials analysed. However, there are examples where efficacy estimates under the two approaches could lead to different scientific conclusions. For example, using published data from a formalin-inactivated hepatitis A vaccine trial, Horne *et al.*[7] calculated a per-protocol efficacy estimate of 1.00 with corresponding 95% confidence interval (CI) of [0.84, 1.00] while the ITT approach yielded an efficacy estimate of 0.81 [0.58, 0.92]. Another example is given by a malaria vaccine trial in children in southern Tanzania for which Alonso *et al.*[46] reported a primary vaccine efficacy estimate of 0.31 [0.00, 0.52] and p-value of 0.046. This primary analysis only considered children of a certain age who received all three doses of the vaccine. Further, a first clinical episode of malaria constituted an endpoint only if it occurred after the third vaccine dose. On the other hand, including all children randomized at first dose and malaria episodes occurring four weeks after the second dose resulted in an estimate of only 0.23 efficacy with 95% CI of $[-0.02, 0.42]$, that is, a nonsignificant result for $\alpha = 0.05$. Even though this second analysis is not strictly ITT according to the definition above, it does illustrate that per-protocol and ITT analysis may give discordant results.

While there is general agreement that some form of an ITT analysis should be performed in clinical trials,[66] the principle is subject to much discussion and alternative approaches remain an active area of research.[67] Ultimately, endpoint data should be collected on all participants whenever feasible regardless of compliance or other circumstances, thus allowing for both ITT and per-protocol analyses.[7] The utility of performing both analyses can be illustrated by considering evaluation of a vaccine's safety profile. Since per-protocol analyses can miss harmful effects of a vaccine in the beginning of the immunization schedule, safety assessment should include an ITT analysis. On the other hand, if a volunteer randomized to control accidently receives vaccine and subsequently develops a serious adverse event, an as-treated analysis including the volunteer in the vaccine arm would provide a more biologically interpretable assessment of safety.

5.5 Case ascertainment and validation

Particular care and consideration should be given to the clinical case definition of the endpoint. For example, it is well known that nonspecific case definitions can lead to attenuated estimates of vaccine efficacy.[8,68,69] Using maximum likelihood based

arguments, Lachenbruch[69] showed that both sensitivity and specificity less than 100% can lead to underestimates of VE, with specificity having a greater impact, especially in rare disease settings. Halloran and Longini[68] suggest that differential endpoint specificity may explain disparate results arising from different randomized controlled trials of comparable live attenuated influenza vaccines. Similarly, Fine and Clarkson[70] note that the variation in estimates of whole-cell pertussis vaccine efficacy may be at least partially attributable to the non-specific nature of clinical criteria in defining pertussis. For example, they note that studies using only bacteriologically confirmed cases yielded higher estimated vaccine efficacies than reported elsewhere.

Given the infeasibility of employing the most specific (and sensitive) case definitions in mass, a few solutions have been suggested to account for misdiagnosed cases in estimating vaccine efficacy. Lachenbruch[69] derived an adjusted VE based on known (or previously estimated) sensitivity and specificity that leads to an increase in the variance estimate of VE and hence requires larger sample sizes to maintain comparable power. Alternatively, Halloran and Longini[68] propose using validation sets wherein the more specific case definition is applied to a randomly selected subset of the participants. The disease or infection rate in the validation sets along with the sampling proportion can be used to adjust the estimate of vaccine efficacy; the mean score method is then employed to obtain confidence intervals.[71] Caution should be exercised when employing this approach in settings where the validation set is not selected by random sampling since further biases can be introduced.[68]

One should also be mindful of the potential for different sensitivity and specificity between vaccine and placebo recipients. For example, Farrington and Miller[27] report some evidence that isolation of *B. pertussis* bacteria is lower in vaccinated individuals exhibiting clinical symptoms compared to placebo recipients, which could result in overestimates of VE.

5.6 Surrogate endpoints

A surrogate endpoint is generally defined as a biomarker that can substitute for a clinically meaningful endpoint for the purpose of comparing specific interventions.[72–74] As mentioned earlier in section 5.3, an example of a surrogate endpoint arises when assessing vaccine effects on post-infection endpoints. In this case, one might consider certain proximal biomarkers in infected participants as surrogate endpoints for the temporally distal disease endpoint of interest. Surrogate endpoints also arise in vaccine efficacy trials within the context of *correlates* of *protective immunity*, that is, biomarkers that are associated with the level of protection from infection or disease due to vaccination.[9,75] Correlates of protection usually entail vaccine-induced immune responses, which, historically, have been defined in terms of antibody titres,[76,77] although current technology allows consideration of cell-mediated, mucosal, and memory-based immune responses as well.[75] In a broader sense, animal model endpoints (as discussed earlier in section 2) might also serve as correlates of protective immunity;[9] for example, attempts have been made to associate efficacy of pertussis vaccines with a mouse model.[78]

From a statistical viewpoint, establishing a putative surrogate endpoint as a valid replacement for the clinical endpoint of interest is challenging since simply demonstrating association is not sufficient. Moreover, we require that the effect of treatment on the

biomarker reliably predicts the treatment's effect on the clinical outcome of interest.[73] In the context of correlates of protective immunity, knowledge of a correlate does not necessarily imply that a vaccine's effect on that particular biomarker will predict efficacy. For example, disease morbidity may be associated with a post-vaccination humoral response even though the true mechanism of protection is cellular-based, simply because humoral and cellular responses are correlated.[75,77] In this case, another vaccine could enhance antibody response levels without improving disease prognosis. Continuing with the post-infection example of HIV, although plasma viral load has been shown to be prognostic for AIDS (and secondary transmission) in natural history studies,[79–82] it remains to be shown that vaccine-induced changes in viral load predict changes in disease progression. A correlate of protection that does indeed predict a vaccine effect on infection or disease is appropriately described as a *surrogate* or *determinant* of protection.

Establishing a surrogate of protective immunity is especially important for future trials where the existence of an efficacious vaccine may result in reduced incidence and preclude the ability to randomize to placebo. Evaluation of combination vaccines, addition of new antigens to or changes in immunization schedules of existing efficacious vaccines, and alteration in manufacturing are all examples in which knowledge of a surrogate of protection is important.[38] Multivalent vaccines present another example where the existence of an established surrogate of protection is critical in that low incidence of certain serotypes may prohibit evaluation of serotype-specific efficacy. For many vaccines, no correlate (much less a surrogate) of protective immunity has been discovered.[9] The elusive search for an immune correlate is illustrated by rotavirus vaccine development, with several trials providing inconsistent results on the correlation of neutralizing antibody titres and serotype-specific vaccine efficacy.[83,84] Even without an established correlate of protection, new combination vaccines, such as acellular pertussis vaccines, have achieved licensure via comparative studies.[75]

Several methods have been proposed for validating a surrogate endpoint.[72,85,86] Prentice[72] offered the first operational criteria, which included that the true endpoint rate be independent of treatment conditional on the history of the surrogate biomarker endpoint. Storsaeter *et al.*[78] employ such an approach in establishing a laboratory (immune) surrogate of protection for pertussis vaccines. Using logistic regression models, they found that given antipertussis antibodies, the risk of developing pertussis does not depend on the type of vaccine received. In some sense, Prentice's definition does not directly apply to validating a correlate of protection within the context of placebo-controlled vaccine trials since vaccine-induced immune responses, by definition, do not exist for the placebo arm. Chan *et al.*[77] make a similar point in an analysis of correlates of protection from a varicella vaccine trial. In general, several proposed methods for validating surrogate endpoints have been met with skepticism[59,74] such that further statistical research is clearly needed. Whatever method is employed, ultimately a clear understanding of the mechanism of protection and disease pathogenesis[75] may likely be required for establishing a valid surrogate endpoint of infection or disease.

5.7 Strain-specific analysis

Many infectious pathogens exhibit considerable genetic and antigenic diversity. Vaccines for such pathogens often contain multiple immunogens, matched to the

major prevalent circulating antigen types (usually serotypes) of the pathogen. In efficacy trials of multivalent vaccines, the primary endpoint has often been disease with any strain homologous (or closely related) to one of the strains contained in the vaccine.[38] For example, this primary endpoint was used in a series of trials of pneumococcal polysaccharide vaccines containing between 6 and 23 serotypes of *Streptococcus pneumonia*.[87,88] The estimate of vaccine efficacy to prevent this endpoint can overestimate field efficacy due to the circulation of heterologous strains that evade vaccine-induced immunity. Accordingly, a secondary analysis of vaccine efficacy to prevent disease with any strain can be useful, and in some trials it is warranted to power the trial for detecting vaccine efficacy against both the 'homologous strain' endpoint and the 'all strain' endpoint. In addition, since the level of protective efficacy can vary by strain, a secondary endpoint in many efficacy trials is strain-specific disease. In this section, we discuss some statistical issues in estimating strain-specific vaccine efficacy. Gilbert *et al.*[89] provide a start toward statistical methods with this purpose.

For vaccines with a known correlate of protective immunity, strain-specific vaccine efficacy can be evaluated quickly in Phase II trials based on an immune response primary endpoint. For example, hemagglutinin-inhibition serological antibody titres to certain influenza strains above a threshold accurately predict protection against influenza infection and illness for certain vaccines, such that vaccine efficacy against a panel of prototype influenza strains can be reliably estimated from a moderate sample of vaccinees.[90,91] In the absence of a known immune correlate, Phase III trials are required for estimating strain-specific vaccine efficacy directly. In such trials, vaccine efficacy against a particular strain has been estimated based on the strain-specific relative cumulative incidence rate (Equation 3) with numerators equal to the number of cases with the particular strain or by a mixed effects logistic regression model[92] for trials with short disease monitoring period, and by a cause-specific Cox proportional hazards model[93] for trials with multiyear follow-up. The former approaches have been used for influenza vaccines, which have shown a pattern of high protective efficacy against influenza strains identical (or nearly so) to a vaccine strain, partial efficacy against strains with minor antigenic changes (for example, due to antigenic drift), and no efficacy against strains with major antigenic changes (for example, due to antigenic shift)[94] (see Sugaya *et al.*[95] for an exception to this pattern). The latter Cox modeling approach has been applied to a large efficacy trial of two cholera vaccines vs. placebo; for each vaccine the method showed superior protection against Classical biotype cholera disease vs. El Tor biotype cholera disease.[96]

An important consideration in assessing strain-specific vaccine efficacy is whether multiple distinct disease events for a subject are counted. For several pathogens, including *Streptococcus pneumoniae*[88] and rotavirus,[97] multiple infection is common and data support that the first infection does not modify the risk of a second infection. For such pathogens, estimation of strain-specific vaccine efficacy is relatively simple: all events are counted and the methods described in section 5.1 can be used with a strain-specific case definition. For other pathogens, however, infection with one strain protects partially or fully against subsequent infection with the same strain or divergent strains (so-called 'interference'); examples include influenza,[92,98] cholera,[99] and HIV.[100] In these cases, partial vaccine efficacy against a strain A can bias the estimate of vaccine efficacy against another strain B, because those 'saved' by vaccine from A disease are

retained in the pool of subjects susceptible to B disease, so that total exposure to B disease is expected to be greater in the vaccine group compared to the placebo group. Competing risks methods that only count first disease episodes have been used when interference is in play. For example, the cause-specific proportional hazards model provides for unbiased estimation of a strain-specific vaccine efficacy parameter under assumptions including that the distribution of exposure to the strain at each follow-up time is the same whether assigned vaccine or placebo.[101] However, substantial vaccine efficacy against a strain A would predict violation of this assumption for another strain B, by creating differential exposure between groups as described above. New statistical methods are needed for correcting strain-specific vaccine efficacy estimates for this potential bias, with important application to HIV vaccine trials.[102,103]

Where vaccine efficacy is inferred to vary by pathogen strain in an efficacy trial, explanatory analyses should be conducted to evaluate if strain-specific vaccine failure is associated with the lack of full immunization, an insufficient immune response to vaccine, or a host characteristic such as age, immune competence, or genotype.[46,104–106] In addition, for assessment of the impact of pathogen type on post-infection endpoints as discussed in section 5.2, the infecting strain can be treated as a covariate, and stratification or regression techniques can be used to estimate strain-specific vaccine efficacy parameters. As discussed earlier, the problems of selection bias and dependent censoring by treatment pose challenges to making these inferences.

5.8 Waning

For most vaccines, durability of efficacy is essential for a vaccination program to control disease in a population,[107,108] although for some vaccines, short-term efficacy is sufficient to control disease (for example, vaccines for travelers and for infants against diseases restricted to early childhood). Vaccine efficacy may wane with time due to declining immunologic memory or to changing antigenicity of the pathogen. If an immune correlate of protection is known, then Phase IV post-licensure studies can be used to track when immune response levels decline below protective levels. (Moulton and Halsey[29] provide a statistical method for this purpose, applied by Mossong *et al.*[109] to demonstrate declining measles antibody titres in children.) There are many examples of vaccines for which evolution of predominant serotypes led to diminished vaccine efficacy, including whole-cell pertussis vaccines[110] and influenza vaccines. The predominant influenza strain changes so rapidly that a new matched immunogen is usually required for the distributed vaccine each year.[111]

For Phase III trials with extensive follow-up, waning has been assessed by estimating the hazard ratio-based vaccine efficacy parameter over time. The methods of Durham *et al.*[112,113] allow for time-varying covariate effects within a Cox-model framework while the method of Gilbert *et al.*[114] provides simultaneous inference on the vaccine efficacy parameter over the duration of the trial. To provide unbiased estimation, these methods rely on an equal exposure assumption in the vaccine and placebo groups at each follow-up time, which is increasingly open to violation as time since randomization increases. In particular, partial vaccine efficacy at intermediate time points can induce bias at future time points by leading to retention of relatively highly exposed vaccine recipients in the risk-set. This problem can be addressed by adjusting for time-dependent covariates predictive of exposure; for example, for sexually transmitted

diseases (STDs) candidate predictor variables include risk behavioral data, STDs from other pathogens, and pathogen-specific CTL responses.[115]

If vaccine efficacy is found to wane in a Phase III trial, then declining immune responses may be suggested as causative if the responses correlate with protection. Changing antigenicity of the pathogen in the geographic region of the trial may be implicated if the infecting strains in the placebo group tend to increase in divergence from vaccine strain(s) over time, and estimates of strain-specific vaccine efficacy decline with the extent of divergence of the exposing strain. A difficult challenge to providing an explanation of waning is that waning is confounded with the mechanism of vaccine protection.[37] To illustrate this problem, note that if a vaccine protects by an all-or-nothing mechanism, then the 'wrong' vaccine efficacy estimator [for example, a proportional-hazards based estimator such as (4)] will be increasingly negatively biased with time; similarly if a vaccine protects by a leaky mechanism, then an attack-rate based estimator such as in expression (3) will tend to be increasingly negatively biased with time. Thus, if the chosen vaccine efficacy estimator is biased for the vaccine efficacy parameter reflecting the 'true' protective mechanism, then efficacy can appear to wane even though it is steady. This problem is difficult to solve because the mechanism of protection is rarely known and is difficult to diagnose.[37] The problem is mitigated for pathogens with relatively low incidence, because in this case both proportional hazards-based and attack-rate-based vaccine efficacy estimators are approximately unbiased regardless of the true protective mechanism.[39,116]

In newer trial designs such as those for HIV vaccines, waning protective efficacy to control viremia or other post-infection endpoints may occur due to development of vaccine resistance, for example due to pathogen mutations in antibody-binding sites, CTL epitopes, or T-helper epitopes.[102,103] Collecting data on genotypic/phenotypic characteristics of the infecting pathogen and on immune responses to autologous pathogen targets over time in infected trial participants will be important for verifying that emergent vaccine resistant mutations lead to loss of protective efficacy.

5.9 Beyond direct efficacy

Heretofore our discussion has focused on endpoints pertaining to the direct effect of a vaccine on the immunized individual. However, from a public health perspective, a vaccine can potentially have other beneficial effects beyond the direct effects of vaccination. For example, vaccinated individuals who become infected may be less likely to transmit the disease to other susceptible individuals. By increasing the degree of a population's immunity to a specific pathogen (or herd immunity[117]), widespread vaccination can also benefit unvaccinated individuals by reducing the probability of contact with an infected individual.[118] Assessing such indirect effects usually entails cluster randomized vaccine trials (CRVTs),[119] with clusters potentially defined by steady sexual partnerships,[120] families, households,[121] schools,[122] clinics,[123] communities,[124] or villages. In simplest form, clusters are randomized to vaccine or placebo wherein all individuals within a cluster receive the same randomization assignment.[122,123,125] Alternatively, a two-step or split-plot randomization scheme[119] could be employed wherein first clusters are randomized not simply to vaccine or placebo, but rather to different vaccination fractions.[118,126] Individual randomization within cluster then follows according to the cluster's randomization assignment. For example,

a cluster might be randomized to a vaccination fraction of 1/3, such that individuals within that cluster are randomized to receive vaccine with probability 1/3 or placebo with probability 2/3. In actuality, some individuals within a cluster might not enrol in the trial, such that a cluster could potentially be partitioned into three subsets: participants randomized to vaccine, participants randomized to placebo, and nonparticipants. Different measures of vaccine efficacy (direct, indirect, total, and overall) can then be estimated via endpoint attack or incidence rate ratios between different subsets of the clusters. These types of efficacy are described briefly below; more thorough treatments are given by Halloran *et al.*[50,126–131]

Direct efficacy refers to the traditional measure of a vaccine's effect as discussed earlier in sections 5.1 and 5.2. Within the context of CRVTs, direct efficacy can be estimated by contrasting attack or incidence rates between participants randomized to vaccine and participants randomized to placebo within a cluster, that is, estimation of direct efficacy within the context of CRVT generally requires a second level of randomization within cluster.[119] Without split-plot randomization, estimates of direct efficacy are subject to bias and should be interpreted with caution.[131] For example, Moulton *et al.*[125] describe design considerations of a CRVT targeting *Streptococcus pneumoniae* in American Indian populations in the southwestern USA. Stratifying by population and geography, approximately 9000 infants within 38 distinct clusters were randomized by cluster to the study or control vaccine. Since within each cluster either all participants were assigned study vaccine or all were assigned the control vaccine, estimation of direct efficacy was not a trial objective given the potential confounding inherent in comparing participants and nonparticipants within a cluster.

Defining other types of efficacy is simplest if we do not consider split-plot randomization. *Indirect efficacy*[125] can be defined as the percent reduction in risk for nonparticipants within a cluster randomized to vaccine compared to nonparticipants within a cluster randomized to placebo. Defined this way, indirect efficacy is less susceptible to the biases of direct efficacy discussed above since comparisons are being made across the same subsets of randomized units (that is, nonparticipants). *Total efficacy*, which incorporates both direct and indirect effects, contrasts endpoint rates in participants within clusters randomized to vaccine and participants within clusters randomized to placebo. *Overall efficacy* compares rates in all individuals within clusters randomized to vaccine and all individuals within clusters randomized to placebo. In other words, total efficacy compares endpoint rates in participants only whereas overall efficacy considers endpoint rates in all individuals within a cluster. Thus, should all individuals in a cluster participate in a CRVT (that is, all individuals in some clusters receive vaccine while all individuals in the remaining clusters receive placebo), the overall and total effects would be the same. Note that these definitions require endpoint ascertainment in nonparticipants; see the last paragraph of this section for further discussion on this point. For split-plot randomized CRVTs, definitions of indirect, total, and overall efficacy can be adapted from Longini *et al.*[126]

More recently, Halloran *et al.*[131] introduce the concept of *epidemic prevention proportion* (EPP) wherein the endpoint is defined as the occurrence of an epidemic in the cluster. The EPP is given by the percent reduction in the probability of an epidemic. As an example, consider the results reported by Bjune *et al.*[122] from a CRVT designed to assess the efficacy of an outer membrane vesicle vaccine against group B meningococcal

disease in which 1335 schools in Norway were randomized to vaccine (690) or placebo (645). The endpoint of interest was an outbreak in the school, defined as at least one case (subsequent cases within a school, of which there was only one, were not counted). Cases of group B meningococcal disease were confirmed in 39 students from schools participating in the study, of which three were excluded due to not meeting the case definition of disease occurring after 14 days post second injection. The remaining 36 cases occurred in 35 schools (11 vaccine, 24 placebo) such that the estimate of efficacy (that is, EPP) was reported as $0.57 = 1 - (11/690)/(24/645)$.

Several statistical issues arise in the context of CRVTs. One of the challenges in the analysis of CRVTs pertains to combining attack rates from different clusters to estimate the EPP, direct, indirect, total, and overall efficacy.[131] For example, summary measures across clusters might entail either taking a ratio of weighted averages of attack rates, or a weighted average of attack rate ratios where the weights are chosen in some optimal fashion. Another consideration arises if individuals are used as the unit of analysis, in which case the intracluster correlation must be accounted for in the analysis. For example, Trach *et al.*[121] report on a trial of a killed, oral cholera vaccine in Vietnam, where all households in the city of Hue were assigned vaccine or placebo. Using an endpoint of cholera requiring inpatient care in a hospital or polyclinic, the estimated protective efficacy was reported in terms of relative attack rates in age-eligible participants from households assigned to vaccine vs. age-eligible participants from households assigned to placebo. Generalized estimating equations were then employed to adjust for possible within-household correlation. Another example is given in Lagos *et al.*,[132] who report results from a large-scale, post-licensure trial of a conjugate *Haemophilus influenzae* type b (Hib) vaccine where 71 urban health centers in Santiago, Chile, served as the unit of randomization. Comparisons of invasive Hib cumulative incidence between infants in health centers randomized to the study vaccine and infants in health centers randomized to the control vaccine were performed using an adaptation of Pearson's chi-square test that appropriately compensates for potential correlation within cluster. Finally, given that CRVTs are usually designed to answer several questions about the efficacy of a candidate vaccine, multiple comparison considerations are again appropriate, especially for trials designed for licensure.[125] Given that the different measures of efficacy discussed here are unlikely to be statistically independent, resampling-based multiple comparison procedures[13] that allow for dependent test statistics might be appropriate.

Although primary endpoints used in CRVTs are typically the same as those in individually- randomized efficacy trials (for example, incidence of infection and/or associated morbidity and mortality) the methods for ascertainment may be vastly different. For large clusters such as communities, the problem of endpoint ascertainment is more similar to that of a disease registry in that the ultimate goal is estimation of population rates. Passive methods of ascertainment (for example, coordinated reporting through existing medical care infrastructures) often provide the foundation for such systems but are not by themselves sufficient for complete case ascertainment. Active methods for case ascertainment may include longitudinal follow-up of defined cohorts within the referent populations, repeated population-based cross-sectional surveys, or some combination of the two. In addition, the nature of important secondary endpoints in CRVTs differs substantially from those in typical individually randomized trials.

In CRVTs, interpretation of observed vaccine effects may depend strongly on assessments of population rates of immigration and emigration, population coverage of vaccination, as well as other economic and operational aspects of the vaccination programme, making accurate measurement of these quantities important.

6 Phase IV studies

For vaccines that prove efficacious in Phase III trials and result in licensure, subsequent Phase IV or post-licensure studies are typically implemented to look at safety and vaccine effectiveness. In general, the term *effectiveness* is used to describe a vaccine's effect in the field post-licensure, whereas *efficacy* pertains to vaccine effect in a well-controlled clinical trial.[12,37,40] An effectiveness trial endpoint typically encompasses all incident cases of infection or disease for any strain of the targeted pathogen occurring during the entire time of follow-up. Overall survival might even be used as an endpoint; for example, Koenig *et al.*[133] report results from a trial showing that measles vaccination had a significant impact on childhood mortality in Bangladesh. Other aspects of Phase IV studies affect endpoint considerations as well. For example, unlike the efficacy trial setting, case validation by collection of appropriate laboratory specimens may not be possible post-licensure, leading to an increased chance of endpoint misclassification. The probability of case classification bias is also increased since post-licensure studies are often not blinded or randomized.[27] For vaccines with a known correlate of protective immunity, endpoints of a Phase IV trial may also include immune response measurements over time, which help inform whether and when booster immunizations are needed. If the pathogen is heterogeneous, the trial endpoints may also encompass monitoring the evolution of the pathogen in a population and expanding the assessment of strain-specific vaccine effectiveness over time and its relationship with vaccine-induced immune responses. Operational aspects of vaccination such as uptake and acceptability might also constitute endpoints in this setting.[40] The reader is referred to Orenstein *et al.*[134] and Clemens *et al.*[40] for further considerations regarding assessment of vaccine effectiveness in the field.

7 Discussion

In this paper we have reviewed endpoint considerations for vaccine trials ranging from preclinical studies in animals to post-licensure field trials. However, the field of vaccine research is vast and several important types of trials and endpoints were not addressed. For example, vaccines motivated by the potential for bioterrorism present distinct challenges[8] since low incidence can often render disease endpoint efficacy trials in humans infeasible. Ideally, correlates of protection would be available in humans, but it may be that animal studies provide the only possible means of evaluating efficacy. In this case, models must be established to bridge results between animal studies and the putative effects of the vaccine in humans. Other scenarios not discussed include trials investigating combination vaccines[135] or vaccines intended to prevent mother-to-child transmission.[48] Endpoints that measure behavioral changes induced by counseling and

other factors that likely will accompany vaccination programs also were not addressed.[47,48,50]

Whatever the setting, the choice of endpoints is one of the most critical aspects in clinical trials. Well-defined endpoints chosen prior to a study's initiation provide guidance in trial design and data analysis while lending scientific and statistical credibility to results.[4] In the context of vaccine trials, measuring clinically significant disease endpoints is critical. While this may prove challenging for diseases with long progression periods, it is essential for understanding clinical vaccine effects and the relevance of surrogate vaccine effects.

Acknowledgement

We would like to thank Jorge Flores for several helpful comments.

References

1 Clemens JD, Stanton BF. Long term evaluation of vaccine efficacy. Methodological issues for Phase III and Phase IV trials. In: Woodrow GC, Levine MM eds. *New generation vaccines*. New York: Marcel Dekker, 1990, pp 51–68.

2 Piantadosi S. *Clinical trials: a methodologic perspective*. New York: Wiley, 1997.

3 Heyse JF. Outcome measures in clinical trials. In: Armitage P, Colton T, eds. *Encyclopedia of biostatistics*. New York: Wiley, 1998, pp 3214–20.

4 Friedman LM, Furberg C, DeMets DL. *Fundamentals of clinical trials*. New York: Springer-Verlag, 1998.

5 Stanley K, Lagakos S. Biostatistical considerations in the design and analysis of AIDS clinical trials. In: Merigan TC, Bartlett JG, Bolognesi D, eds. *Textbook of AIDS medicine*, 2nd edn. Baltimore: Williams and Wilkins, 1999, pp 807–14.

6 Ellenberg SS, Dixon DO. Statistical issues in designing clinical trials of AIDS treatments and vaccines. *Journal of Statistical Planning and Inference* 1994; **42**: 123–35.

7 Horne AD, Lachenbruch PA, Goldenthal KL. Intent-to-treat analysis and preventive vaccine efficacy. *Vaccine* 2001; **19**: 319–26.

8 Horne AD, Clifford J, Goldenthal KL, Kleppinger C, Lachenbruch PA. Efficacy and safety of preventive vaccines: FDA evaluation. In *Proceedings of the Annual Meeting of the American Statistical Association*, Atlanta, August 5–9, 2001.

9 Clements-Mann ML. Lessons for AIDS vaccine development from non-AIDS vaccines. *AIDS Research and Human Retroviruses* 1998; **14**(Suppl 3): S197–203.

10 Almeida R, Norrish A, Levick M, Vetrie D, Freeman T, Vilo J, Ivens A, Lange U, Stober C, McCann S, Blackwell J. From genomes to vaccines: *Leishmania* as a model. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 2002; **357**: 5–11.

11 De Groot AS, Sbai H, Aubin CS, McMurry J, Martin W. Immuno-informatics: mining genomes for vaccine components. *Immunology and Cell Biology* 2002; **80**: 255–69.

12 Strathdee SA, Loughlin AM. Vaccines – past, present, and future. In: Nelson KE, Williams CM, Graham NMH, Masters CF, eds. *Infectious disease epidemiology: theory and practice*. Gaithersburg, MD: Aspen Publishers, 2001, pp 255–82.

13 Westfall PH, Young SS. *Resampling-based multiple testing: examples and methods for p-value adjustment*. New York: Wiley, 1993.

14 Mehrotra DV, Heyse JF. Multiplicity considerations in clinical safety analyses. In *Proceedings of the Annual Meeting of the American Statistical Association*, Atlanta, USA, 5–9 August 2001.

15 ICH Harmonised Tripartite Guidelines: Statistical Principles for Clinical Trials. *Federal Register* 1998; **63**(179): 49583–98.

16 Belshe RB, Stevens C, Gorse GJ, Buchbinder S, Weinhold K, Sheppard H, Stablein D, Self S, McNamara J, Frey S, Flores J, Excler JL, Klein M, El Habib R, Duliege A, Harro C, Corey L, Keefer M, Mulligan M, Wright P, Celum C, Judson F, Mayer K, McKirnan D, Marmor M, Woody G, for the National Institute of Allergy and Infectious Diseases AIDS Vaccine Evaluation Group, and HIV Network for Prevention Trials (HIVNET). Safety and immunogenicity of a canarypox-vectored human immunodeficiency virus type 1 vaccine with or without gp120: A phase 2 study in higher- and lower-risk volunteers. *Journal of Infectious Diseases* 2001; **183**: 1343–52.

17 Clements-Mann ML, Weinhold K, Matthews TJ, Graham BS, Gorse GJ, Keefer MC, McElrath MJ, Hsieh R, Mestecky J, Zolla-Pazner S, Mascola J, Schwartz D, Siliciano R, Corey L, Wright PF, Belshe R, Dolin R, Jackson S, Xu S, Fast P, Walker MC, Stablein D, Excler J, Tartaglia J, Duliege A, Sinangil F, Paoletti E, and the NIAID AIDS Vaccine Evaluation Group. Immune responses to human immunodeficiency virus (HIV) type 1 induced by canarypox expressing HIV-1MN gp120, HIV-1SF2 recombinant gp120, or both vaccines in seronegative adults. *Journal of Infectious Diseases* 1998; **177**: 1230–46.

18 Corey L, McElrath MJ, Weinhold K, Matthews T, Stablein D, Graham B, Keefer M, Schwartz D, Gorse G, and the AIDS Vaccine Evaluation Group. Cytotoxic T cell and neutralizing antibody responses to human immunodeficiency virus type 1 envelope with a combination vaccine regimen. *Journal of Infectious Diseases* 1998; **177**: 301–309.

19 Evans TG, Keefer MC, Weinhold KJ, Wolff M, Monteffori D, Gorse GJ, Graham BS, McElrath MJ, Clements-Mann ML, Mulligan MJ, Fast P, Walker MC, Excler J, Duliege A, Tartaglia J, and the NIAID AIDS Vaccine Evaluation Group. A canarypox vaccine expressing multiple human immunodeficiency virus type 1 genes given alone or with rgp120 elicits broad and durable CD8+ cytotoxic T lymphocyte responses in seronegative volunteers. *Journal of Infectious Diseases* 1999; **180**: 290–98.

20 AIDS Vaccine Evaluation Group 022 Protocol Team. Cellular and humoral immune responses to a canarypox vaccine containing human immunodeficiency virus type 1 env, gag, and pro in combination with rgp120. *Journal of Infectious Diseases* 2001; **183**: 563–70.

21 Russell N, Hudgens MG, Ha R, Havenar-Daughton C, McElrath MJ. Moving to HIV-1 vaccine efficacy trials: defining T cell responses as potential correlates of immunity. *Journal of Infectious Diseases* 2003; **187**: 226–42.

22 Kahn P. NIH drops plans for Phase III trial. International AIDS Vaccine Initiative Report, 6, 2002.

23 Self SG, Hudgens MG, Chiu Y, Rossini AJ. Statistical methods for evaluating ELISPOT assays in HIV vaccine trials. *AIDS Vaccine 2001*, Philadelphia, September 2001. Poster no. 36.

24 Broman K, Speed T, Tigges M. Estimation of antigen-responsive T cell frequencies in PBMC from human subjects. *Journal of Immunological Methods* 1998; **98**(2): 119–32.

25 Herrington DA. Initial clinical evaluation of new vaccine candidates. Phase I and II trials of safety, immunogenicity, and preliminary efficacy. In: Woodrow GC, Levine MM, eds. *New generation vaccines*. New York: Marcel Dekker, 1990, pp 43–49.

26 Rao MR, Blackwelder WC, Troendle JF, Naficy AB, Clemens JD. Sample size determination for phase II studies of new vaccines. *Vaccine* 2002; **20**: 3364–69.

27 Farrington P, Miller E. Clinical trials. In: Robinson A, Farrar G, Wiblin C, eds. *Methods in molecular medicine: vaccine protocols*. Totowa: Humana Press, 1996, pp 251–68.

28 Carey VJ, Baker CJ, Platt R. Bayesian inference on protective antibody levels using case-control data. *Biometrics* 2001; **57**(1): 135–42.

29 Moulton LH, Halsey NA. A mixed gamma model for regression analyses of quantitative assay data. *Vaccine* 1996; **14**: 1154–58.

30 Moulton LH, Holt EA, Job JS, Halsey NA. Percentile regression analysis of correlated antibody responses. Statistics in Medicine 1995; **14**: 777–87.

31 Moulton LH, Halsey NA. A mixture model with detection limits for regression analyses of antibody response to vaccine. *Biometrics* 1995; **51**: 1570–78.

32 Hudgens MG. Estimating cumulative probabilities from incomplete longitudinal binary responses with application to HIV vaccine trials. *Statistics in Medicine* 2003; **22**: 463–79.

33 Flores J, Perez-Schael I, Gonzalez M, Garcia D, Perez M, Daoud N, Cunto W, Chanock RM, Kapikian AZ. Protection against rotavirus diarrhoea by rhesus rotavirus vaccine in Venezuelan children. *Lancet* 1987; **1**: 882–84.

34 Flores J, Perez-Schael I, Blanco M, Rojas AM, Alfonso E, Crespo I, Cunto W, Pittman A, Kapikian AZ. Reactogenicity and immunogenicity of a high titered rhesus rotavirus based quadrivalent rotavirus vaccine. *Journal of Clinical Microbiology* 1994; **31**: 2439–45.

35 Cochran WG, Cox GM. *Experimental designs*, 2nd edn (Classics Edition). New York: Wiley, 1992.

36 Levine MM, Kaper JB, Herrington D, Ketley J, Losonsky G, Tacket CO, Tall B, Cryz S. Progress in the development of an HIV-1 vaccine. *Lancet* 1988; **2**(8609): 467–70.

37 Farrington CP. Communicable diseases. In: Armitage P, Colton T, eds. *Encyclopedia of biostatistics*. New York: Wiley, 1998; pp 795–815.

38 Goldenthal KL, Falk LA, Geber A. Prelicensure evaluation of combination vaccines. *Clinical Infectious Diseases* 2001; **33**(Suppl 4): S267–73.

39 Smith PG, Rodrigues LC, Fine PEM. Assessment of the protective efficacy of vaccines against common diseases using case–control and cohort studies. *International Journal of Epidemiology* 1984; **13**: 87–93.

40 Clemens J, Brenner R, Rao M, Tafari N, Lowe C. Evaluating new vaccines for developing countries. Efficacy or effectiveness? *Journal of the American Medical Association* 1996; **275**: 390–97.

41 Chick SE, Barth-Jones DC, Koopman JS. Bias reduction for risk ratio and vaccine effect estimators. *Statistics in Medicine* 2001; **20**: 1609–24.

42 Szmuness W, Stevens CE, Zang EA, Harley EJ, Kellner Z. A controlled clinical trial of the efficacy of the hepatitis B vaccine Heptavax B: a final report. *Hepatology* 1981; **1**: 377–85.

43 Kaplan EL, Meier P. Nonparametric estimation from incomplete data. *Journal of the American Statistical Association* 1958; **53**: 457–81.

44 Peto R. Empirical survival curves for interval censored data. *Applied Statistics* 1973; **22**: 86–91.

45 Yip PSF, Chen Q. A partial likelihood estimator of vaccine efficacy. *The Australian and New Zealand Journal of Statistics* 2000; **42**(3): 367–74.

46 Alonso PL, Smith T, Armstrong Schellenberg JRM, Masanja H, Mwankusye S, Urassa H, Bastos de Azevedo I, Chongela J, Kobero S, Menendez C, Hurt N, Thomas MC, Lyimo E, Weiss NA, Hayes R, Kitua AY, Lopez MC, Kilama WL, Teuscher T, Tanner M. Randomised trial of efficacy of SPf66 vaccine against *Plasmodium falciparum* malaria in southern Tanzania. *Lancet* 1994; **344**: 1175–81.

47 Schaper C, Fleming TR, Self SG, Rida WN. Statistical issues in the design of HIV vaccine trials. *Annual Review of Public Health* 1995; **16**: 1–22.

48 Gilbert PB. Some statistical issues in the design of HIV-1 vaccine and treatment trials. *Statistical Methods in Medical Research* 2000; **9**(3): 207–29.

49 Haber M, Longini IM, Halloran ME. Measures of the effects of vaccination in a randomly mixing population. *International Journal of Epidemiology* 1991; **20**: 300–10.

50 Halloran ME, Longini IM, Struchiner CJ. Design and interpretation of vaccine field studies. *Epidemiological Reviews* 1999; **21**: 73–88.

51 Gilbert PB, Bosch R, Hudgens MG. Sensitivity analysis for the assessment of causal vaccine effects on viral load in HIV vaccine trials. *Biometrics* 2003; **59**: 531–41.

52 Hudgens MG, Hoering A, Self SG. On the analysis of viral load endpoints in HIV vaccine trials. *Statistics in Medicine* 2003; **22**: 2281–98.

53 Francis DP, Gregory T, McElrath MJ, Belshe RB, Gorse GJ, Migasena S, Kitayaporn D, Pitisuttitham P, Matthews T, Schwartz DH, Berman PW. Advancing AIDSVAX to phase 3: safety, immunogenicity, and plans for phase 3. *AIDS Research and Humun Retroviruses* 1998; **14**(Suppl 3): S325–31.

54 Fine PEM. Implications of different study designs for the evaluation of acellular pertussis vaccines. *Developments in Biological Standardization* 1997; **89**: 123–33.

55 Gilbert PB, DeGruttola V, Hudgens MG, Self SG, Hammer SM, Corey L. What constitutes efficacy for an human immunodeficiency virus vaccine that ameliorates viremia: issues involving surrogate end points in phase 3 trials. *Journal of Infectious Diseases* 2003; **188**: 179–93.

56 Rotnitzky A, Robins JM, Scharfstein DO. Semiparametric regression for repeated outcomes with nonignorable nonresponse. *Journal of the American Statistical Association* 1998; **93**: 1321–39.

57 Chang MN, Guess HA, Heyse JF. Reduction in burden of illness: a new efficacy measure for prevention trials. *Statistics in Medicine* 1994; **13**: 1807–14.

58 Vesikari T, Rautanen T, Varis T, Beards GM, and Kapikian AZ. Rhesus *Rotavirus* candidate vaccine: clinical trial in children vaccinated between 2 and 5 months of age. *American Journal of Diseases of Children* 1990; **144**: 285–89.

59 Frangakis CE, Rubin DB. Principal stratification in causal inference. *Biometrics* 2002; **58**: 21–29.

60 Kliks SC, Nisalak A, Brandt WE, Wahl L, Burke DS. Antibody-dependent enhancement of dengue virus growth in human monocytes as a risk factor for dengue hemorrhagic fever. *American Journal of Tropical Medicine and Hygiene* 1989; **40**: 444–51.

61 Nabel GJ. Challenges and opportunities for development of an AIDS vaccine. *Nature* 2001; **410**: 1002–1007.

62 Burke DS. Human HIV vaccine trials: does antibody-dependent enhancement pose a genuine risk? *Perspectives in Biology and Medicine* 1992; **35**(4): 511–30.

63 Mascola JR, Mathieson BJ, Zack PM, Walker MC, Halstead SB, Burke DS. Summary report: workshop on the potential risks of antibody-dependent enhancement in human HIV vaccine trials. *AIDS Research and Human Retroviruses* 1993; **9**(12): 1175–84.

64 Prentice RL. On the role and analysis of secondary outcomes in clinical trials. *Statistics in Medicine* 1989; **8**: 431–40.

65 Lachin JM. Statistical considerations in the intent-to-treat principle. *Controlled Clinical Trials* 2000; **21**: 167–89.

66 Ellenberg JN. Intention-to-treat analysis. In: Armitage P, Colton T, eds. *Encyclopedia of biostatistics*. New York: Wiley, 1998, pp 2056–60.

67 Goetghebeur E, Loeys T. Beyond intent to treat. *Epidemiologic Reviews* 2002, **24**: 85–90.

68 Halloran ME, Longini IM. Using validation sets for outcomes and exposure to infection in vaccine field studies. *American Journal of Epidemiology* 2001; **154**: 391–98.

69 Lachenbruch PA. Sensitivity, specificity, and vaccine efficacy. *Controlled Clinical Trials* 1998; **19**: 569–74.

70 Fine PEM, Clarkson JA. Reflections on the efficacy of pertussis vaccines. *Reviews of Infectious Diseases* 1987; **9**: 866–83.

71 Pepe MS, Reilly M, Fleming TR. Auxiliary outcome data and the mean score method. *Journal of Statistical Planning and Inference* 1994; **42**: 137–60.

72 Prentice RL. Surrogate endpoints in clinical trials: Definition and operational criteria. *Statistics in Medicine* 1989; **8**: 431–40.

73 Fleming TR, DeMets DL. Surrogate end points in clinical trials: are we being misled? *Annals of Internal Medicine* 1996; **125**: 605–13.

74 DeGruttola VG, Clax P, DeMets DL, Downing, GJ, Ellenberg SS, Friedman L, Gail MH, Prentice R, Wittes J, Zeger SL. Considerations in the evaluation of surrogate endpoints in clinical trials: summary of a National Institutes of Health workshop. *Controlled Clinical Trials* 2001; **22**: 485–502.

75 Edwards KM. Development, acceptance, and use of immunologic correlates of protection in monitoring the effectiveness of combination vaccines. *Clinical Infectious Diseases* 2001; **33**: S274–77.

76 Siber GR. Methods for estimating serological correlates of protection. *Developments in Biological Standardization* 1997; **89**: 283–96.

77 Chan ISF, Li S, Matthews H, Chan C, Vessey R, Sadoff J, Heyse J. Use of statistical models for evaluating antibody response as a correlate of protection against varicella. *Statistics in Medicine* 2002; **21**: 3411–30.

78 Storsaeter J, Hallander HO, Gustafsson L, Olin P. Levels of anti-pertussis antibodies related to protection after households exposure to *Bordetella pertussis*. *Vaccine* 1998; **16**: 1907–16.
79 Mellors JW, Kingsley LA, Rinaldo Jr CR, Todd JA, Kokka RP, Gupta P. Quantitation of HIV-1 RNA in plasma predicts outcome after seroconversion. *Annals of Internal Medicine* 1995; **122**: 573–79.
80 Katzenstein TL, Pedersen C, Nielsen C, Lundgren JD, Jakobsen PH, Gerstoft J. Longitudinal serum HIV RNA quantification: correlation to viral phenotype at seroconversion and clinical outcome. *AIDS* 1996; **10**(2): 167–73.
81 Pedersen C, Katzenstein T, Nielsen C, Lundgren JD, Gerstoft G. Prognostic value of serum HIV-RNA levels at virologic steady state after seroconversion: relation to CD4 cell count and clinical course of primary infection. *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology* 1997; **16**: 93–99.
82 Quinn TC, Wawer MJ, Sewankambo N, Serwadda D, Li C, Wabwire-Mangen F, Meehan MO, Lutalo T, Gray RH. Viral load and heterosexual transmission of human immunodeficiency virus type 1. *New England Journal of Medicine* 2000; **342**: 921–29.
83 Ward RL, Bernstein DI, Young EC, Sherwood JR, Knowlton DR, Schiff GM. Human rotavirus studies in volunteers: determination of infectious dose and serological response to infection. *Journal of Infectious Diseases* 1986; **154**: 871–80.
84 Lanata CF, Black RE, del Aguila R, Gil A, Verastegui H, Gerna G, Flores J, Kapikian AZ, Andre FE. Protection of Peruvian children against rotavirus diarrhea of specific serotypes of one, two, or three doses of the RIT 4237 attenuated bovine rotavirus vaccine. *Journal of Infectious Diseases* 1989; **159**: 452–59.
85 Buyse M, Molenberghs G. Criteria for the validation of surrogate endpoints in randomized experiments. *Biometrics* 1998; **54**: 1014–29.
86 Begg CB, Leung DHY. On the use of surrogate end points in randomized trials (pkg: P15-28). *Journal of the Royal Statistical Society, Series A, General* 2000; **163**(1): 15–24.
87 Smit P, Oberholzer D, Hayden-Smith S, Koornhof HJ, Hilleman MR. Protective efficacy of pneumococcal polysaccharide vaccines. *Journal of the American Medical Association* 1977; **238**: 2613–16.
88 Bruyn GRW, van Furth R. Pneumococcal polysaccharide vaccines: indications, efficacy and recommendations. *European Journal of Clinical and Microbiological Infectious Diseases* 1991; **10**: 897–910.
89 Gilbert PB, Self SG, Ashby MA. Statistical methods for assessing differential vaccine protection against human immunodeficiency virus types. *Biometrics* 1998; **54**: 799–814.
90 Masurel N, Laufer J. A one year study of trivalent influenza vaccines in primed and unprimed volunteers: immunogenicity, clinical reactions, and protection. *Journal of Hygiene* 1984; **92**: 263–75.
91 Longini IM, Koopman JS, Haber M, Cotsonis GA. Statistical inference for infectious diseases: risk-specific household and community transmission parameters. *American Journal of Epidemiology* 1988; **128**: 845–59.
92 Longini IM, Halloran ME, Nizam A, Wolff M, Mendelman PM, Fast PE, Belshe RB. Estimation of the efficacy of live, attenuated influenza vaccine from a two-year, multicenter vaccine trial: implications for influenza epidemic control. *Vaccine* 2000; **18**: 1902–909.
93 Prentice RL, Kalbeisch JD, Peterson AV, Flournoy N, Farewell VT, Breslow NB. The analysis of failure times in the presence of competing risks. *Biometrics* 1978; **34**: 541–54.
94 Gross PA, Hermogenes AW, Sack HS, Lau J, Levandowski RA. The efficacy of influenza vaccine in elderly persons. A meta-analysis and review of the literature. *Annals of Internal Medicine* 1995; **123**: 518–27.
95 Sugaya N, Nerome K, Ishida M, Matsumoto M, Mitamura K, Nirasawa M. Efficacy of inactivated vaccine in preventing antigenically drifted influenza type A and well-matched type B. *Journal of the American Medical Association* 1994; **272**: 1122–26.
96 Clemens JD, Sack DA, Harris JR, Van Loon F, Chakraborty J, Ahmed F, Rao MR, Khan MR, Yunus MD, Huda N, Stanton BF, Kay BA, Walter S, Eeckels R, Svennerholm AM, Holmgren J. Field trial of oral cholera vaccines in Bangladesh: results from three-year follow-up. *Lancet* 1990; **335**: 270–73.

97 Kapikian AZ, Hoshino Y, Chanock RM, Perez-Schael I. Efficacy of quadrivalent rhesus rotavirus-based human rotavirus vaccine aimed at preventing severe rotavirus diarrhea in infants and young children. *Journal of Infectious Diseases* 1996; **174**(Suppl 1): S65–72.

98 Andreasen V, Lin J, Levin SA. The dynamics of cocirculating influenza strains conferring partial cross-immunity. *Journal of Mathematical Biology* 1997; **35**: 825–42.

Q2 99 Clemens JD, van Loon F, Sack DA *et al.* Biotype as determinant of natural immunising effect of cholera. *Lancet* 1991; **337**: 883–84.

Q2 100 Travers K, Mboup S, Marlink R, Gueye-Nidaye A, Siby T, Thior I, Traore I, Dieng-Sarr A, Sankale JL, Mullins C *et al.* Natural protection against HIV-1 infection provided by HIV-2. *Science* 1995; **268**: 1612–15.

101 Gilbert PB. Comparison of competing risks failure time methods and time-independent methods for assessing strain variations in vaccine protection. *Statistics in Medicine* 2000; **19**(22): 3065–86.

102 UNAIDS Report from a meeting of the WHO-UNAIDS Vaccine Advisory Committee Geneva, 21–23 February 2000. Approaches to the development of broadly protective HIV vaccines: challenges posed by genetic, biological and antigenic variability of HIV-1. *AIDS* 2001; **15**: W1–25.

103 Lukashov VV, Goudsmit J, Paxton WA. The genetic diversity of HIV-1 and its implications for vaccine development. In: Wong-Staal F, Gallo RC, eds. *AIDS vaccine research*. New York: Marcel Dekker, 2002, pp 93–120.

104 van Loon F. Cholera: developments in prevention and cure. *Tropical and Geographical Medicine* 1993; **45**: 269–73.

105 Hogan CM, Hammer SM. Host determinants in HIV infection and disease Part II: genetic factors and implications for antiretroviral therapeutics. *Annals of Internal Medicine* 2001; **134**: 978–96.

106 Gilbert P, Self S, Rao M, Naficy A, Clemens J. Sieve analysis: methods for assessing how vaccine efficacy depends on genotypic and phenotypic pathogen variation from vaccine trial data. *Journal of Clinical Epidemiology* 2001; **54**: 68–85.

107 Anderson RM, Garnett GP. Low-efficacy HIV vaccines: potential for community-based intervention programmes. *Lancet* 1996; **348**: 1010–13.

108 Anderson RM, Swinton J, Garnett GP. Potential impact of low-efficacy HIV-1 vaccines in populations with high rates of infection. *Proceedings of the Royal Society of London B* 1995; **261**: 1010–13.

109 Mossong J, O'Callaghan CJ, Rantnam S. Modelling antibody response to measles vaccine and subsequent waning of immunity in a low exposure population. *Vaccine* 2001; **19**: 523–29.

110 PHLS Whooping-Cough Committee and Working Party. Efficacy of whooping-cough vaccines used in the United Kingdom before 1968. *British Medical Journal* 1973; 259–63. Q3

111 World Health Organization. Recommended composition of influenza virus vaccines for use in the 2003–2004 influenza season. *Weekly Epidemiological Record* 2003; **78**: 57–64.

112 Durham LK, Longini IM, Halloran ME, Clemens JD, Nizam A, Rao M. Estimation of vaccine efficacy in the presence of waning: application to cholera vaccines. *American Journal of Epidemiology* 1998; **147**: 948–59.

113 Durham LK, Halloran ME, Longini IM, Manatunga AK. Comparison of two smoothing methods for exploring waning vaccine effects. *Applied Statistics* 1999; **48**: 395–407.

114 Gilbert PB, Wei LJ, Kosorok MR, Clemens JD. Simultaneous inference on the contrast of two hazard functions with censored observations. *Biometrics* 2002; **58**: 773–80.

115 Dorrell L, Dong T, Ogg SG, Lister S, McAdam S, Rostron T, Conlon C, McMichael AJ, Rowland-Jones SL. Distinct recognition of non-clade B human immunodeficiency virus type 1 epitopes by cytotoxic T lymphocytes generated from donors infected in Africa. *Journal of Virology* 1999; **73**: 1708–14.

116 Halloran ME, Haber M, Longini IM. Interpretation and estimation of vaccine efficacy under heterogeneity. *American Journal of Epidemiology* 1992; **136**: 328–43.

117 Anderson RM, May RM. *Infectious diseases in humans: dynamics and control*. Oxford: Oxford University Press, 1991.

118 Longini IM, Sagatelian K, Rida WN, Halloran ME. Optimal vaccine trial design when estimating vaccine efficacy for susceptibility and infectiousness from multiple populations (corr: 1999v18 p890). *Statistics in Medicine* 1998; **17**: 1121–36.

119 Hayes RJ, Alexander NDE, Bennett S, Cousens SN. Design and analysis issues in cluster-randomized trials of interventions against infectious diseases. *Statistical Methods in Medical Research* 2000; **9**(2): 95–116.

120 Longini IM, Hudgens MG, Halloran ME. Estimation of vaccine efficacy for both susceptibility to infection and reduction in infectiousness for prophylactic HIV vaccines with partner augmentation. In: Kaplan E, Brookmeyer R, eds. *The quantitative evaluation of HIV prevention programs*. New Haven: Yale University Press, 2002, pp 241–59.

121 Trach DD, Clemens JD, Ke NT, Thuy HT, Son ND, Canh DG, Hang PVD, Rao MR. Field trial of a locally produced, killed, oral cholera vaccine in Vietnam. *Lancet* 1997; **349**: 231–35.

122 Bjune G, Hoiby EA, Gronnesby JK, Arnesen O, Holst-Fredriksen J, Halstensen A, Holten E, Lindbak A, Nokleby H, Rosenqvist E, Solberg LK, Closs O, Eng J, Froholm LO, Lystad A, Bakketeig LS, Hareide B. Effect of outer membrane vesicle vaccine against group B meningococcal disease in Norway. *Lancet* 1991; **338**: 1093–96.

123 Jaffar S, Leach A, Hall AJ, Obaro S, McAdam KPWJ, Smith PG, Greenwood BM. Preparation for a pneumococcal vaccine trial in The Gambia: individual or community randomization. *Vaccine* 1999; **18**: 633–40.

Q4 124 Pollock TM. Trials of prophylactic agents for the control of communicable diseases. *World Health Organization Monograph Series* 1966; 52.

125 Moulton LH, O'Brien KL, Kohberger R, Chang I, Reid R, Weatherholtz R, Hackell JG, Siber GR, Santosham M. Design of a group-randomized *Streptococcus pneumoniae* vaccine trial. *Controlled Clinical Trials* 2001; **22**: 438–52.

126 Longini IM, Halloran ME, Nizam A. Model-based estimation of vaccine effects from community vaccine trials. *Statistics in Medicine* 2002; **21**(4): 481–95.

127 Halloran ME, Struchiner CJ. Study designs for dependent happenings. *Epidemiology* 1991; **2**: 331–38.

128 Halloran ME, Struchiner CJ. Causal inference in infectious diseases. *Epidemiology* 1995; **6**: 142–51.

129 Halloran ME, Struchiner CJ, Longini IM. Study designs for evaluating different efficacy and effectiveness aspects of vaccines. *American Journal of Epidemiology* 1997; **146**: 789–803.

130 Halloran ME. Vaccine studies. In: Armitage P, Colton T, eds. *Encyclopedia of biostatistics*. New York: Wiley, 1998, pp 4687–94.

131 Halloran ME, Longini IM, Cowart DM, Nizam A. Community interventions and the epidemic prevention potential. *Vaccine* 2002; **20**: 3254–62.

132 Lagos R, Horwitz I, Toro J, Martin OS, Abrego P, Bustamante C, Wasserman SS, Levine OS, Levine MM. Large scale, postlicensure, selective vaccination of Chilean infants with PRP-T conjugate vaccine: practicality and effectiveness in preventing invasive *Haemophilus influenzae* type b infections. *Pediatric Infectious Diseases* 1996; **15**: 216–22.

133 Koenig MA, Khan MA, Wojtyniak B, Clemens JD, Chakraborty J, Fauveau V, Phillips JF, Akbar J, Barua US. Impact of measles vaccination on childhood mortality in rural Bangladesh. *Bulletin of the World Health Organization* 1990; **68**: 441–47.

134 Orenstein WA, Berner RH, Hinman AR. Assessing vaccine efficacy in the field. *Epidemiologic Reviews* 1988; **10**: 212–41.

135 Potema AS, Myers MG, Breiman RF. Challenges in the development, licensure, and use of combination vaccines. *Clinical Infectious Diseases* 2001; **33**: S261–66.