MAJOR ARTICLE

A Framework for Assessing Immunological Correlates of Protection in Vaccine Trials

Li Qin,[1,3] Peter B. Gilbert,[1,3] Lawrence Corey,[2,4,5,6] M. Juliana McElrath,[2,4,5] and Steven G. Self[1,3]

[1]Statistical Center for HIV/AIDS Research and Prevention and [2]Program in Infectious Diseases, Fred Hutchinson Cancer Research Center, and Departments of [3]Biostatistics, [4]Laboratory Medicine, [5]Medicine, and [6]Microbiology, University of Washington, Seattle

(See the editorial commentary by Sadoff and Wittes, on pages 1279–81.)

A central goal of vaccine research is to identify a vaccine-induced immune response that predicts protection from infection or disease. The term "correlate of protection" has been used to refer to at least 3 distinct concepts, which has resulted in confusion surrounding this topic. We propose precise definitions of these different concepts of immune correlates, using the nomenclature "correlate of risk," "level 1 surrogate of protection," and "level 2 surrogate of protection." We suggest a general framework for assessing these 3 levels of immune correlates in vaccine efficacy trials. To demonstrate the proposed principles, we analyze data from a 1943 influenza vaccine field trial, supporting Weiss strain A–specific antibody titers as a level 1 surrogate of protection. Other real and simulated examples are also discussed.

A central goal of vaccine research is to identify a vaccine-induced immune response that predicts protection from infection or disease [1–4]. Such responses are mainly used to predict a vaccine's protective effect in a new setting, for which vaccine efficacy (VE) is not directly observed. For example, immune responses may be used to predict protection induced by a vaccine across vaccine lots, human populations, viral populations, and even species. If these predictions are reliable, then use of such immune correlates provides an efficient way to guide the development, evaluation, and utilization of vaccines. However, empirically validating such predictions is challenging.

Despite the importance of identifying immunological correlates of protection (CoPs) and the extensive literature reporting attempts to find them, the methodology available for their quantitative assessment is limited [1, 5–7]. Moreover, at least 3 different conceptual definitions have been implicitly used for CoPs, which has created confusion and controversy in the literature. These different concepts may be organized in a hierarchy that is related to the strength of the empirical basis for the CoP's validity as a predictor. Typically, the confusion results from a claim for validity of a CoP at a conceptual level that is higher than what the empirical validation supports. We see a need to clarify the CoP terminology and to build a rigorous framework for assessing immunological CoPs.

Here, we distinguish 3 distinct concepts, each having been described as a CoP, and map them to concepts described in the literature on surrogate end points [8–16]. We provide an ordering of these concepts in terms of their proximity to the ultimate definition of a correlate as a predictor of protection for new settings and describe the data requirements for rigorous validation of an immunological measurement at each level. The evaluation approaches are illustrated by use of past vaccine trials, with a 1943 influenza vaccine field trial of a trivalent vaccine as our central example [17]. We selected influenza vaccination as a prototype for discussion because its potential effectiveness appears to be the most likely scenario for many candidate vaccines in clinical trials, such as those for HIV-1 and herpes

Received 30 December 2006; accepted 16 April 2007; electronically published 2 October 2007.

Potential conflicts of interest: none reported.

Presented in part: HIV Vaccine Trials Network Full Group Meeting, Washington, DC, 23–24 May 2006; HIV Vaccine Trials Network Conference, Seattle, 16–18 October 2006.

Financial support: National Institute of Allergy and Infectious Diseases, National Institutes of Health (grants U01 AI068635, R37 AI029168, and R01 AI054165-04).

Reprints or correspondence: Dr. Li Qin, Statistical Center for HIV/AIDS Research and Prevention, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave. N, LE-400, Seattle, WA (lqin@scharp.org).

The Journal of Infectious Diseases 2007; 196:1304–12

0022-1899/2007/19609-0008$15.00
DOI: 10.1086/522428

simplex virus–1, and for newly emerging immunotherapeutic vaccines for cancer. The literature is replete with articles that use antibodies to the influenza hemagglutinin protein as a surrogate of VE. We work through some original data that developed this concept and assess titers of antibody to Weiss strain A and to PR8 strain A at the 3 levels of immune correlates.

Table 1 defines the 3-tiered framework for evaluating immune correlates. We now provide details for each tier.

CORRELATE OF RISK (COR)

The primary clinical end point used in VE trials is pathogen-specific morbidity and/or mortality [2]. In some settings, other end points might be used, such as infection or postinfection viremia in HIV vaccine studies [8]. We refer to an immunological measurement that predicts a clinical end point in some population as a CoR.

The CoR concept has been used in different contexts. In observational studies, immune responses of exposed HIV-seronegative individuals have been referred to as CoRs [22]. In VE trials, acute immune responses to the vaccine that correlate with the rate of clinical end point may be termed CoRs [23]. To validate an immunological measurement as a CoR, there must be a source of variability in the measurements, and an association must be observed between these measurements and the pathogen-specific clinical end point. As discussed below, for some infections for which multiple reexposures to the pathogen can occur, an immunological measurement may have substantial variability in unvaccinated persons, so that it can be evaluated as a CoR in nonvaccinees as well as in vaccinees. If study participants have no prior exposure to the pathogen, however, the immune response to the vaccine may be negative for (almost) all nonvaccinees, precluding its evaluation as a CoR in nonvaccinees.

We use published data from the 1943 influenza vaccine study [17] to demonstrate the assessment of a potential CoR. In this study, the names of 1776 male participants were alphabetized. Every other participant was inoculated either with 1 mL of a trivalent vaccine containing Weiss strain A, PR8 strain A, and Lee strain B antigens or with the subcutaneous control. The primary end point was hospitalization due to influenza. Titers of strain-specific antibody to the vaccine were evaluated as CoRs of strain-specific influenza virus infection, defined as incidence of hospitalization with a respiratory illness plus the identification of a particular strain of influenza virus in throat culture. Figure 1 shows distributions of the $\log_2$ strain-specific serum antibody titers. Results from a logistic regression model fitted to the data are summarized in table 2. For the control group, the titers of antibody to Weiss strain A are highly inversely associated with infection/hospitalization incidence ($P < .0001$), showing them to be a strong CoR, whereas the titers of antibody to PR8 strain A are weakly associated ($P = .08$) and, hence, are a poor CoR. Subsequent studies of influenza virus infection demonstrated an association between strain-specific antibody titers and infection or morbidity, substantiating this immunological measurement as a CoR [24].

SURROGATE OF PROTECTION (SOP)

A SoP is a CoR that reliably predicts a vaccine's level of protective efficacy on the basis of contrasts in the vaccinated and unvaccinated groups' immunological measurements. Because there are different data requirements for validating a SoP for predicting VE for the same setting (vaccine, population, etc.) of the trial than for predicting efficacy for different settings not considered in the trial, we distinguish SoPs at 2 levels for these 2 cases, naming them "level 1 SoPs" and "level 2 SoPs." We discuss their evaluation in the following sections.

A CoR fails to be a SoP if it cannot adequately explain the vaccine's effect on the clinical end point. For example, a recent efficacy trial of an HIV vaccine identified a CoR that was not a SoP. The levels of antibody blocking of gp120 binding to soluble CD4 inversely correlated with the HIV infection rate in the vaccinated group, identifying a CoR, but the absence of protective efficacy against HIV infection strongly supports that the CoR is not a SoP [23]. See the literature on surrogate end points [10] for discussions about how a CoR can fail to be a SoP.

Different measures of VE have been defined [25]. For a typical efficacy trial, VE is the percent reduction in the risk of clinically significant infection for the vaccinated group versus that for the control group:

$$\mathrm{VE} = \left[1 - \frac{\Pr(\text{infection}|\text{vaccine})}{\Pr(\text{infection}|\text{control})}\right] \times 100\% \ .$$

Before evaluating an immunological CoR as a potential SoP, there needs to be evidence that $\mathrm{VE} > 0$. In the 1943 influenza field trial [17], the Weiss strain A–specific infection incidence was 2.25% for vaccinees and 8.45% for control subjects, and the PR8 strain A–specific incidence was 2.25% for vaccinees and 8.22% for control subjects. The estimated VE was 73% for each strain, with a 95% confidence interval of 57%–84% for Weiss strain A and of 55%–83% for PR8 strain A. These results justify assessing each antibody variable as a potential SoP.

LEVEL 1 SOP

We consider 2 analytic approaches to the evaluation of a level 1 SoP on the basis of data from a single large VE trial. The first approach identifies a SoP as a surrogate end point that satisfies the Prentice criterion [16], an empirical criterion that can be directly assessed with the data available from a standard efficacy trial. The Prentice criterion requires that the observed protective effect of the vaccine can be completely explained in

Table 1. Terminology for immunological measurements as 3 levels of immune correlates—correlate of risk (CoR), level 1 surrogate of protection (SoP), and level 2 SoP.

Term	Definition	Framework for empirical assessment	Data analytic methods
CoR	An immunological measurement that correlates with the rate or level of a study end point used to measure VE in a defined population	Efficacy trials or observational studies	Regression models
Level 1 SoP	An immunological measurement that is a CoR within a defined population of vaccinees and is predictive of VE in the same setting as the trial; validation entails showing either level 1 SoP^S or level 1 SoP^P		
Level 1 SoP^S	The relationship between the immunological measurement and the risk of the study end point is the same in vaccinees and nonvaccinees	Single large VE trial	Statistical surrogate framework [16]
Level 1 SoP^P	The criterion defined by Frangakis and Rubin [18] and by P.B.G. and M. Hudgens (unpublished data): (1) groups of subjects with no or the lowest vaccine effect on the immune response have no VE and (2) groups of subjects with a sufficiently large vaccine effect on the immune response have positive VE	Single large VE trial	Principal surrogate framework [18–21]
Level 2 SoP	An immunological measurement that is a level 1 SoP and that is predictive of VE in different settings (e.g., across vaccine lots, across human populations, across viral populations, across species)	Multiple VE trials and/or postlicensure studies	Meta-analysis [11–15]

NOTE. SoP^P, SoP principal; SoP^S, SoP statistical; VE, vaccine efficacy.

a statistical model by the immunological measurements. The Prentice surrogate definition is most useful for immunological measurements that have substantial variability among control subjects, because this provides a basis for comparing the immune-response effect on risk in both the vaccinated and unvaccinated groups.

A second approach for assessing a level 1 SoP is based on the principal surrogate framework of causal inference [18–21].

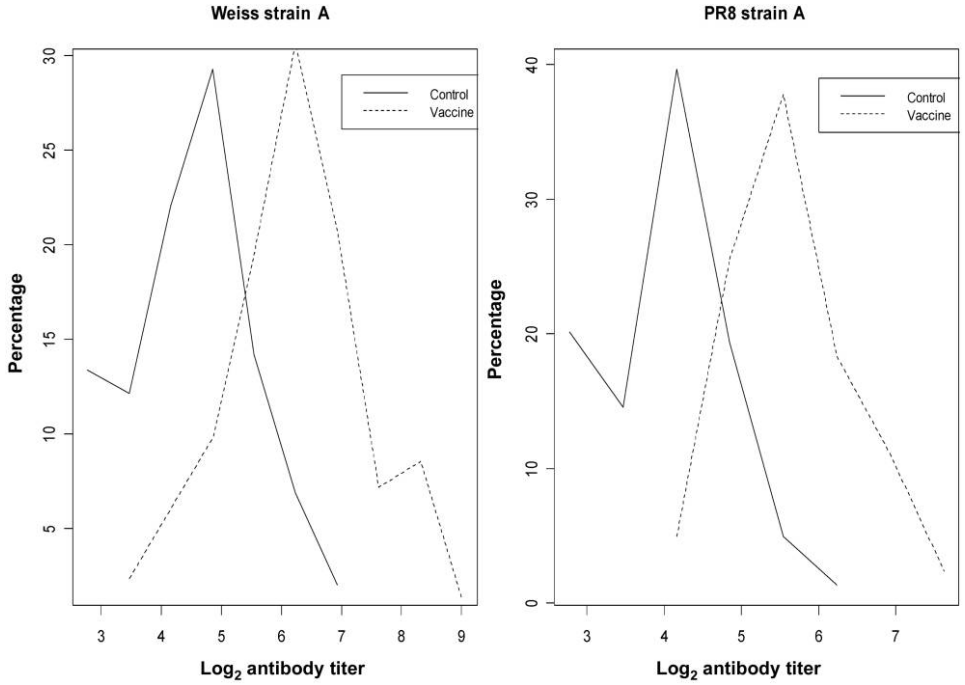

Figure 1. Distribution of strain-specific $\log_2$ neutralizing antibody titers

Table 2. Logistic regression fit to strain-specific log_2 neutralizing-antibody titers.

Category, parameter	Weiss strain A		PR8 strain A	
	Estimated coefficient (SE)	*P*	Estimated coefficient (SE)	*P*
Control group only				
Intercept	1.80 (0.54)	.001	−1.37 (0.59)	.021
Log_2 titer	−1.03 (0.14)	<.0001	−0.27 (0.15)	.077
Vaccine and control groups				
Intercept	1.62 (0.45)	.0003	−1.27 (0.53)	.0172
Log_2 titer	−0.98 (0.12)	<.0001	−0.29 (0.13)	.0310
Vaccination status	−0.33 (0.32)	.3068	−0.89 (0.34)	.0085

In this framework, potential outcomes are imagined that represent what would occur to an individual under each potential condition of randomization to the vaccine and control groups. An immunological measurement is considered to be a level 1 SoP if (1) groups of vaccinees with absent or the lowest response levels have a risk equal to that had they not been vaccinated and (2) groups of vaccinees with sufficiently high immune response levels have a risk lower than that had they not been vaccinated. Because this definition compares risk among groups with identical characteristics except for vaccination status, any difference is directly attributable to vaccine and, thus, is a causal effect [19].

The 2 types of level 1 SoPs are referred to as a "SoP statistical" (SoP^S) and a "SoP principal" (SoP^P), following terms coined in the statistical literature [18]. Discussion of SoP assessment within each framework follows.

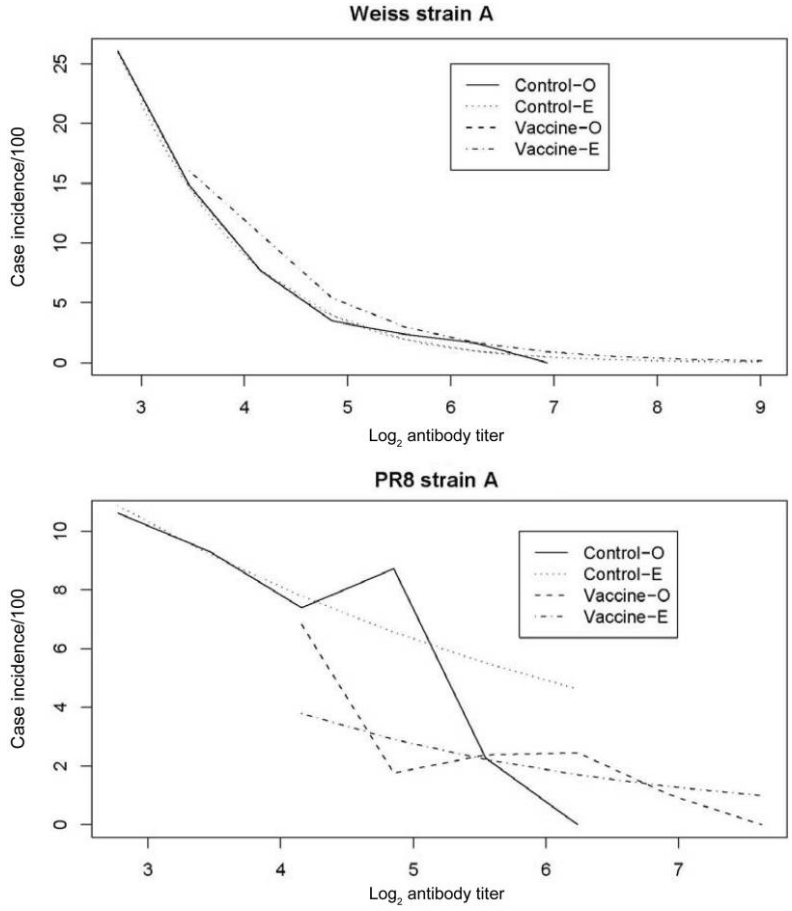

Figure 2. Observed (O) and predicted (expected [E]) infection incidences for the vaccine and control groups

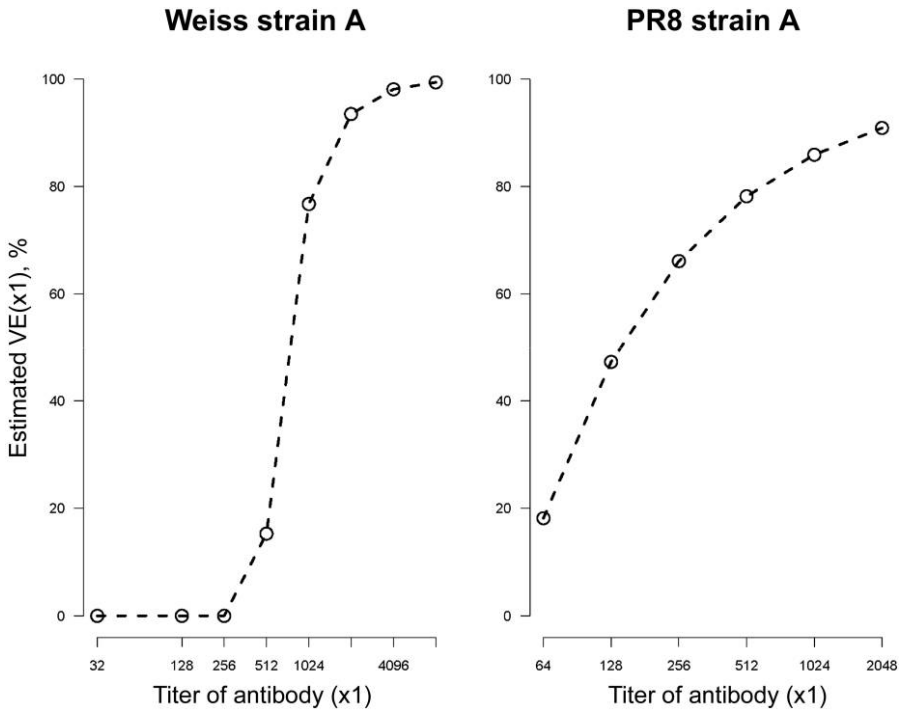

Figure 3. Estimated VE(x1) for $\log_2$ titers of neutralizing antibody to Weiss strain A and to PR8 strain A

LEVEL 1 SOP^S

The data requirements for assessing a potential SoP^S are difficult to achieve, particularly when surrogacy is imperfect [10–13]. Imperfect surrogates are likely for newer vaccine types that are directed at inducing T cell responses, for which the employed assays measure only a few of the potential myriad number of functions that vaccine- or pathogen-specific T cells can produce. However, if an excellent SoP^S exists, then it is possible to identify it in a single large trial.

Figure 2 displays observed and predicted strain-specific infection incidences from logistic regression fits of the $\log_2$ titers of antibody to Weiss strain A and to PR8 strain A in the 1943 influenza vaccine trial [17]. The figure shows that, after controlling for titers of antibody to Weiss strain A, the risk of infection is virtually the same among the vaccinated and unvaccinated groups (for $\log_2$ titer, $P < .0001$; for vaccination group, $P > .1$), supporting these titers as a SoP^S. Further support derives from the observation that the predicted VE based on titers of antibody to Weiss strain A is close to the directly observed VE (82% and 73%, respectively). Significantly, this might represent the first example of a biomarker outcome that has been empirically validated to satisfy the Prentice criterion as a perfect surrogate end point.

In contrast, figure 2 shows that, after controlling for titers of antibody to PR8 strain A, there remain differences in infection risk between the groups ($P = .008$). Moreover, the predicted VE based on these antibody titers is only 33%, compared with the observed 73%. These results support that the protection against PR8 strain A influenza is conferred through mechanisms not fully captured in the assay for neutralizing antibody to PR8 strain A. Therefore, titers for PR8 strain A appear to be an imperfect level 1 SoP^S.

LEVEL 1 SOP^P

A SoP^S is defined purely in terms of statistical and observable associations. However, validation of a SoP^S is based on comparing risk between groups that are selected after randomization by their immune response values. Thus, the statistical surrogate framework has been criticized for its susceptibility to postrandomization selection bias, which may make this framework misleading for making reliable predictions [18]. To address this problem, a new framework for evaluating surrogates has been developed on the basis of causal effects [18, 21] (P.B.G. and M. Hudgens, unpublished data; L.Q., P.B.G., D. Follmann, and D. Li, unpublished data).

To assess whether an immunological measurement is a SoP^P, we need to study how VE varies over groups defined by fixed values of the immune response if assigned vaccine, or X(1). That is, we need to estimate

$$\mathrm{VE}(x1) = \left\{1 - \frac{\Pr[\mathrm{INF}^V | \mathrm{X}(1) = x1]}{\Pr[\mathrm{INF}^C | \mathrm{X}(1) = x1]}\right\} \times 100\% \ ,$$

where INF^V is infection if assigned vaccine and INF^C is infection

Table 3. Summary of the 29 influenza trials [26].

	Cases/no. of	
Year	Vaccinees	Nonvaccinees
1966	7/171	7/79
1966	8/159	8/79
1968	52/465	16/59
1968	91/471	17/59
1969	25/1254	42/413
1969	206/933	227/841
1969	91/881	95/521
1969	166/1030	95/521
1969	27/187	5/25
1972	16/384	35/340
1969	68/1947	59/977
1969	65/1961	59/978
1976	75/116	68/109
1983	31/121	24/59
1986	21/91	19/88
1987	75/878	46/439
1988	373/1060	193/532
1989	276/1126	180/563
1990	229/1016	119/508
1995	249/409	287/416
1994	16/77	3/12
1984	75/300	84/298
1985	111/457	56/241
1986	209/577	99/253
1987	200/723	72/217
1988	202/789	40/145
1998	161/576	132/544
1999	82/582	128/596
1997	86/294	98/299

if assigned control. This VE parameter is interpreted as the percent reduction in the risk for groups of vaccinees with immune response x1, compared with that if they had not been vaccinated.

To estimate VE(x1), one must predict X(1), the immune response that an unvaccinated subject would have had if vaccinated. Follmann [21] introduced 2 approaches to predicting X(1): (1) baseline irrelevant predictor, which includes incorporating a baseline variable that is measured in both the vaccinated and unvaccinated groups, that correlates with the immune response of interest, and that does not predict clinical risk after accounting for X(1); and (2) closeout placebo vaccination, which includes vaccinating a sample of control subjects uninfected at the end of the trial and measuring their immune response—X(1)—to vaccine. Statistical methods have been developed that use these approaches to estimate VE(x1), and simulation studies have demonstrated their utility [21] (P.B.G. and M. Hudgens, unpublished data; L.Q., P.B.G., D. Follmann, and D. Li, unpublished data).

We here demonstrate the evaluation of a SoP^P by use of the influenza example [17], with X(1) being the log titer of antibody to Weiss strain A or to PR8 strain A if assigned vaccine. A baseline variable predicting X(1) was not measured in this trial, nor was closeout placebo vaccination performed, so we will use a different approach for predicting X(1) for nonvaccinees. Because data suggest that prevaccination titers of antibody to influenza virus are inversely correlated with postvaccination titers in adults [24], we will make an antiequipercentile assumption. Specifically, we assume that the X(1)s of nonvaccinees are in the inverse ranking order as the titers actually measured for these nonvaccinees. For Weiss strain A, the predicted X(1) given the observed titer x1 of a nonvaccinee is as follows: for an x1 of 16, an X(1) of 8192 is predicted; for 32, 4096; for 64, 2048; for 128, 1024; for 256, 512; for 512, 256; and for 1024, 32 or 128 (each with a probability of 0.5). For PR8 strain A, the predictions are as follows: for 16, 2048; for 32, 1024; for 64, 512; for 128, 256; for 256, 128; and for 512, 64. Logistic regression models were used to estimate the probabilities of infection at each level $X(1) = x1$ observed in the vaccine group. Figure 3 displays the resulting estimates of VE(x1). The results support that Weiss strain A titers have high value as a SoP^P, because the estimated VE(x1) is zero if the vaccine produces low titers—$X(1) < 512$—and increases to 1.0 if it produces titers $\geqslant 1024$. The results also suggest that PR8 strain A titers have partial value as a SoP^P, because the estimated VE(x1) increases from 0.2 to 0.85 for x1 increasing from 64 to 2048. This imputation-based assessment relies strongly on the assumptions we have made, and trials with either the baseline irrelevant predictor or closeout placebo vaccination strategy could potentially evaluate a surrogate with more realistic assumptions.

LEVEL 2 SOP

The ultimate goal of immune correlate evaluation is to identify an immunological measurement that reliably predicts VE across settings different from those studied in an efficacy trial. Such a correlate can facilitate rapid and objective assessment of vaccine prototypes and their refinements and can guide the expansion of vaccination to novel populations—for example, to immunocompromised patients. We refer to such a "cross-predictive" immune correlate as a level 2 SoP.

Because a level 2 SoP is a group-level predictor of vaccine effects on risk across different settings, meta-analysis [11–15] is suitable for evaluating a level 2 SoP. The meta-analytic unit and the goals of the prediction are the key elements of the assessment. For example, to predict VE against a new viral strain, the meta-analytic unit should be a circulating viral strain, and strain-specific assessments of vaccine immunogenicity and efficacy are required. These assessments can be performed with a very large phase 3 trial or across multiple phase 2b/3/4 efficacy

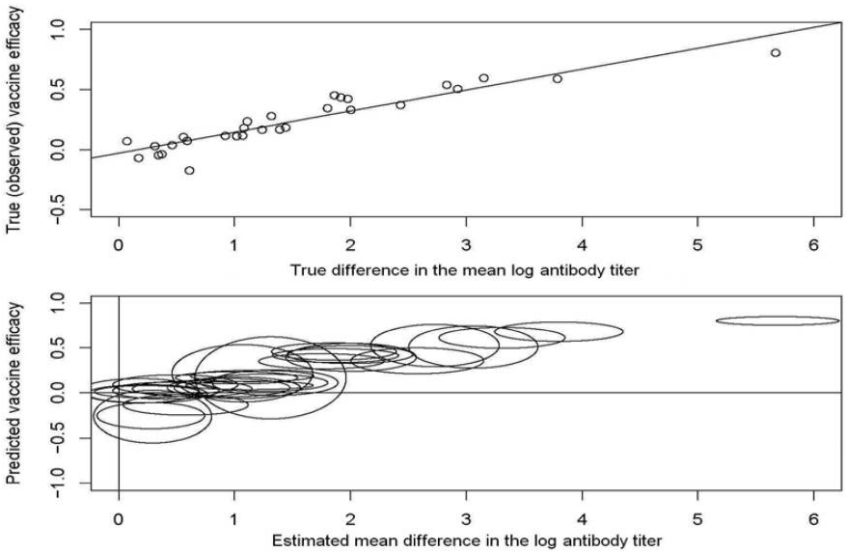

Figure 4. True and estimated clinical effects (vaccine efficacy) and surrogate effects (antibody levels) of the vaccine. Ellipses represent 95% confidence regions associated with the estimated results from each study. See table 3 for a summary of the 29 influenza trials [26].

trials. The observed relationship between the estimated VEs and the differences in immune responses between vaccinees and nonvaccinees provide the basis for predicting VE in a new setting given the observed immune responses in that setting.

We illustrate a hypothetical meta-analysis to assess whether the identified influenza virus strain–specific level 1 SoP is useful for predicting the vaccine's effect on emerging viral strains. Because the influenza study [17] measured only 2 strain-specific antibody titers, we simulated 29 randomized clinical trials of influenza vaccines, with a distinct circulating strain in each trial. We used the sample sizes and estimated VEs from clinically confirmed cases of influenza in real trials (selected from table 1 in Villari et al. [26]). All trials of parainfluenza virus vaccine with at least 3 influenza cases in the control group were included. Table 3 and figure 4 summarizes the 29 simulated trials and shows the association between the observed and predicted clinical and immunological effects. The association conforms to the relationship between the true parameters. The meta-analysis approach is very data intensive and may not always be feasible. Moreover, with a genetically variable pathogen such as influenza virus or HIV-1, the ability to develop large data sets that support precise evaluation for many pathogen strains is difficult. Inferences from meta-analyses always involve some extrapolation, and as such incorporating information on the biological mechanism of protection is important for building the credibility of a level 2 SoP.

For 3 vaccines, the appendix provides brief case studies of the knowledge level about the immune correlates at the 3 levels. Our search of the literature revealed some articles that implicitly evaluated a level 1 SoP^S by use of the Prentice criterion, as discussed here (including examples 2 and 3 in the appendix). However, no articles were found that evaluated a level 1 SoP^P, which requires augmented data collection.

DISCUSSION

The assessment of immune correlates is a key issue in vaccine trials. An immune correlate can be used for guiding vaccine development and refinement, for predicting VE in different settings, and for guiding vaccination policies and regulatory decisions. In this article, we have proposed a general framework for assessing an immunological measurement as a CoR, a level 1 SoP, and, ultimately, a level 2 SoP.

The proposed framework is organized in a logical hierarchy reflecting increased difficulty in achieving different levels of assessment. The assessment of a CoR is relatively straightforward and is achievable with standard efficacy trial designs, and that of a level 1 SoP is difficult, with pros and cons for both the statistical and principal surrogate evaluation frameworks. Direct evaluation of a level 2 SoP must be based on large-scale efficacy trials and/or postlicensure studies that provide ample statistical power to evaluate VE across several different settings.

Selection of immunological measurements to be assessed at the 3 levels of immune correlates is largely driven by knowledge of underlying biological processes and the plausibility of proposed mechanisms for protecting against infection or disease. However, to make reliable assessments, measurement error properties of the immunological assays must be addressed. For highly precise assays such as those used in the influenza trial example [17], this is not an issue. However, in other trials, such as current HIV vaccine trials, the primary assays (including T cell assays) may produce noisy measurements. Such measure-

ment error can strongly attenuate statistical power for detecting immune correlates at any of the 3 levels [33]. As such, it is important to assess measurement error and components of variation of immunological assays and to integrate this information into the design of VE trials, to ensure that they are adequately powered to evaluate immune correlates.

In fact, studies designed solely to detect VE may be very underpowered to assess immune correlates. Vaccine trials to assess immune correlates should at a minimum be powered to detect a CoR, and, where possible, a level 1 SoP. In particular, investigators might consider collecting additional data on baseline risk factors and predictors of immune responses to the vaccine and powering the trials to detect a level 1 SoP^P. Another strategy to consider is to vaccinate a sample of control group participants after they complete follow-up and measure the potential level 1 SoP^P. Augmenting trial designs with extra data collection holds potential for improving assessments of correlates, compared with what can be achieved with standard trial designs. Last, standardization in immunological measurements and efficacy end points across VE trials will be a particularly important programmatic goal to the extent that it will enable assessment of a level 2 SoP via meta-analysis.

APPENDIX

ILLUSTRATION OF HOW 3 VACCINES FIT INTO THE 3 LEVELS OF IMMUNE CORRELATES

Example 1: hepatitis B vaccine—CoRs validated as a level 2 SoP or as a level 1 but not a level 2 SoP. In several studies of young and older adults, immunocompetent subjects, and immunocompromised subjects, no recipients of Recombivax HB (Merck) vaccine who maintained anti–hepatitis B surface antigen (anti-HBs) concentrations $\geqslant$10 mIU/mL were observed to acquire clinically significant hepatitis B virus (HBV) infection (reviewed in [27]). This suggests that at-exposure levels of anti-HBs $\geqslant$10 mIU/mL perfectly predict protection across diverse populations. In contrast, anti-HBs concentrations immediately following the Recombivax regimen are supported as a level 1 SoP but not a level 2 SoP. Immunocompetent vaccinees with initial postvaccination concentrations $\geqslant$10 mIU/mL have been observed to be perfectly and durably protected regardless of whether the concentrations wane to <10 mIU/mL, whereas clinically significant breakthrough HBV infections have been observed in immunocompromised vaccinees whose concentrations waned to <10 mIU/mL [27].

Example 2: acellular pertussis vaccine—CoRs validated or invalidated as a level 1 SoP. Several pediatric studies of acellular pertussis vaccines have identified CoRs for pertussis disease, including postvaccination levels of antibody to pertactin, fimbriae, pertussis toxin, and filamentous hemagglutinin [28, 29]. Almost all investigations of immune correlates reported in the literature have been done at the CoR level, and there is much uncertainty about the value of the different serological measurements as SoPs [30]. As an exception, Storsaeter et al. [28] applied regression models that supported anti-pertactin, anti-fimbriae, and anti–pertussis toxin antibody levels as a joint level 1 SoP^S. The field of pertussis vaccine development may benefit from undertaking further assessments of level 1 and level 2 SoPs, which may entail novel trial designs or data collection.

Example 3:* Haemophilus influenzae *type b vaccine—CoR invalidated as a level 1 SoP. Kayhty et al. [31] measured serum antibodies to capsular polysaccharide of *H. influenzae* type b vaccine in 514 children and evaluated the distributions of these concentrations with respect to (1) the age-specific incidence of meningitis due to *H. influenzae* type b in Finland from 1975 to 1981 and (2) the age-specific VE measured in a large efficacy trial [32]. Kayhty et al. [31] noted that the relationship between meningitis incidence and antibody concentrations was different in vaccinees and nonvaccinees, which invalidates these concentrations as a level 1 SoP^S. Specifically, concentrations $\geqslant$0.15 μg/mL predicted a very low incidence in nonvaccinees, whereas concentrations $\geqslant$1.0 μg/mL were needed to predict a very low incidence in vaccinees.

References

1. Fauci AS, Haynes BF, Pantaleo G. Toward an understanding of the correlates of protective immunity to HIV infection. Science **1996**; 271: 324–8.
2. Clements-Mann ML. Lessons for AIDS vaccine development from non-AIDS vaccines. AIDS Res Hum Retroviruses **1998**; 14(Suppl 3): S197–203.
3. Fauci AS. The AIDS epidemic—considerations for the 21st century. N Engl J Med **1999**; 341:1046–50.
4. Burton DR, Desrosiers RC, Doms RW, et al. HIV vaccine design and the neutralizing antibody problem. Nat Immunol **2004**; 5:233–6.
5. Siber GR. Methods for estimating serological correlates of protection. Dev Biol Stand **1997**; 89:283–96.
6. Chan ISF, Shu L, Matthews H, et al. Use of statistical models for evaluating antibody response as a correlate of protection against varicella. Stat Med **2002**; 21:3411–30.
7. Dunning AJ. A model for immunological correlates of protection. Stat Med **2006**; 25:1485–97.
8. Gilbert PB, DeGruttola V, Hudgens M, Self S, Hammer S, Corey L. What constitutes efficacy for an HIV vaccine that ameliorates viremia: issues involving surrogate end points in phase 3 trials. J Infect Dis **2003**; 188:179–93.
9. Fleming TR. Surrogate markers in AIDS and cancer trials. Stat Med **1994**; 13:1423–35.
10. DeMets DL, Fleming TR. Surrogate endpoints in clinical trials: are we being misled? Ann Intern Med **1996**; 125:605–13.
11. De Gruttola VG, Clax P, DeMets DL, et al. Considerations in the evaluation of surrogate endpoints in clinical trials. Control Clin Trials **2001**; 22:485–502.
12. Hughes MD. Evaluating surrogate endpoints. Control Clin Trials **2002**; 23:703–7.

13. Molenberghs G, Buyse M, Geys H, Renard D, Burzykowski T, Alonso A. Statistical challenges in the evaluation of surrogate endpoints in randomized trials. Control Clin Trials **2002**; 23:607–25.
14. Daniels MJ, Hughes MD. Meta-analysis for the evaluation of potential surrogate markers. Stat Med **1997**; 16:1965–82.
15. Gail MH, Pfeiffer R, Van Houwelingen HC, Carroll RJ. On meta-analytic assessment of surrogate outcomes. Biostatistics **2000**; 1:231–46.
16. Prentice RL. Surrogate endpoints in clinical trials: definition and operational criteria. Stat Med **1989**; 8:431–40.
17. Salk JE, Menke WJ Jr, Francis T Jr. A clinical, epidemiological and immunological evaluation of vaccination against epidemic influenza. Am J Hyg **1945**; 42:57–93.
18. Frangakis CE, Rubin DB. Principal stratification in causal inference. Biometrics **2002**; 58:21–9.
19. Rubin DB. Causal inference using potential outcomes: design, modeling, decisions. J Am Stat Assoc **2005**; 100:322–31.
20. Holland P. Statistics and causal inference. J Am Stat Assoc **1986**; 81: 945–61.
21. Follmann D. Augmented designs to assess immune response in vaccine trials. Biometrics **2006**; 62:1161–70.
22. Zhu T, Corey L, Hwangbo Y, et al. Persistence of extraordinarily low levels of genetically homogeneous human immunodeficiency virus type 1 in exposed seronegative individuals. J Virol **2003**; 77:6108–16.
23. Gilbert PB, Peterson ML, Follmann D, et al. Correlation between immunologic responses to a recombinant glycoprotein 120 vaccine and incidence of HIV-1 infection in a phase 3 HIV-1 preventive vaccine trial. J Infect Dis **2005**; 191:666–77.
24. Gorse GJ, O'Connor TZ, Newman FK, et al. Immunity to influenza in older adults with chronic obstructive pulmonary disease. J Infect Dis **2004**; 190:11–9.
25. Halloran ME, Longini I, Struchiner C. Design and interpretation of vaccine field studies. Epidemiol Rev **1999**; 21:73–88.
26. Villari P, Manzoli L, Boccia A. Methodological quality of studies and patients age as major sources of variation in efficacy estimates of influenza vaccination in healthy adults: a meta-analysis. Vaccine **2004**; 22:3475–86.
27. Mast EE, Weinbaum CM, Fiore AE, et al. A comprehensive immunization strategy to eliminate transmission of hepatitis B virus infection in the United States: recommendations of the Advisory Committee on Immunization Practices (ACIP). Part II: immunization of adults. MMWR Recomm Rep **2006**; 55(RR-16):1–33.
28. Storsaeter J, Hallander HO, Gustafsson L, Olin P. Levels of anti-pertussis antibodies related to protection after household exposure to *Bordetella pertussis*. Vaccine **1998**; 16:1907–16.
29. Taranger J, Trollfors B, Lagergard T, et al. Correlation between pertussis toxin IgG antibodies in postvaccination sera and subsequent protection against pertussis. J Infect Dis **2000**; 181:1010–3.
30. Millen SH, Bernstein DI, Connelly B, Ward JI, Chang SJ, Weiss AA. Antibody-mediated neutralization of pertussis toxin-induced mitogenicity of human peripheral blood mononuclear cells. Infect Immun **2004**; 72:615–20.
31. Kayhty H, Peltola H, Karanko V, Makela PH. The protective level of serum antibodies to the capsular polysaccharide of *Haemophilus influenzae* type b. J Infect Dis **1977**; 136(Suppl):S43.
32. Peytola H, Kayhty H, Sivonen A, Makela H. *Haemophilus influenzae* type b capsular polysaccharide vaccine in children: a double-blind field study of 100,000 vaccinees 3 months to 5 years of age in Finland. Pediatrics **1977**; 60:730–7.
33. Carrol RJ, Ruppert D, Stefanski LA. Measurement error in nonlinear models. Boca Raton, FL: Chapman & Hall, **1998**.