Evaluating Immune Correlates in HIV Type 1 Vaccine Efficacy Trials: What RV144 May Provide

Morgane Rolland¹ and Peter Gilbert²

Abstract

Since the RV144 vaccine combination showed efficacy in a Phase III trial, it provides an opportunity to generate hypotheses about the immune responses necessary for protection against HIV-1 infection, and these results could help devise vaccine candidates with higher efficacy. Here we describe how researchers can determine the correlates of immune protection for an HIV/AIDS vaccine, particularly in the context of the RV144 trial, and we discuss the terminology used to describe correlates and surrogates.

Introduction

THE POSITIVE RESULTS OF THE RV144 trial provided impetus to perform a variety of assays aimed at deciphering the biological mechanisms behind the modest protection conferred by the vaccine combination. A large collaborative network of investigators was formed to perform pilot immunogenicity studies in preparation for a case-control study designed to identify correlates of immune protection. Identifying one or more correlates would help the iterative development of vaccine candidates. It would allow for smaller trials of shorter duration; it could be used to inform regulatory decisions, and would facilitate the extrapolation of efficacy to new trial settings.

Given how research questions are framed in HIV articles or funding requests, establishing immune correlates is a priority for HIV research, yet the statistical considerations that bound the identification of a correlate are not necessarily emphasized. The RV144 case-control study may yield one or multiple correlates of the rate of HIV-1 infection in the vaccine group, but is unlikely to provide a surrogate endpoint for HIV-1 infection, given that more data are needed to assess surrogates requiring augmented trial designs. Here, we address statistical issues for the assessment of correlates and surrogates in the RV144 pilot and case-control studies.

Correlate vs. Surrogate

For the RV144 trial, we define a correlate as a measured vaccine-induced immune response (e.g., estimated IC₅₀ based on a dilution series) that is associated with the rate of HIV-1 infection in the vaccine group, while a surrogate endpoint is a correlate that reliably predicts the level of protection from

infection. Standard efficacy trials only permit the assessment of correlates; then, further analyses are needed to evaluate their surrogate value. Surrogates are a more reliable basis for the development of future vaccines, and they can be used as study endpoints for follow-up Phase I/II vaccine trials and, ultimately, in bridging studies of a licensed vaccine. Correlation is necessary but not sufficient for a measured immune response to be a surrogate; moreover, there is a continuum in the “surrogate value” of a correlate: some may be worthless, some are essentially perfect, and others have intermediate predictive value. For example, a correlate is worthless as a surrogate if the protection does not differ with the vaccine-induced immune response, while a correlate is a perfect surrogate if protection is nil for subjects with none or little of the correlated vaccine-induced immune response and is >90% for subjects with an immune response above a threshold.

A statistical framework for evaluating biomarkers as surrogate endpoints in clinical trials was pioneered in 1989 by Ross Prentice,¹ who defined a surrogate endpoint as “a response variable for which a test of the null hypothesis of no relationship to the treatment groups under comparison is also a valid test of the corresponding null hypothesis based on the true endpoint.” For efficacy trials, this definition means that the vaccine affects the immune response only if the overall vaccine efficacy is positive. Prentice’s definition of a surrogate requires two key conditions: the surrogate needs to be a correlate of infection (as noted above) and to capture all of the vaccine effect (heuristically, “full-mediation” of the protective efficacy). In 2002, Frangakis and Rubin² pointed out a limitation of the Prentice criteria: a validated Prentice surrogate may fail to reliably predict vaccine efficacy if the analysis of full mediation does not account for every subject

¹MHRP:HJF, Rockville, Maryland.

²Vaccine and Infectious Disease Institute, Fred Hutchinson Cancer Research Center, Seattle, Washington.

characteristic that is predictive of both HIV-1 infection and of the potential surrogate; such a factor could easily arise, due to host genetics for example. Thus, they proposed a new surrogate definition to avoid this pitfall.

More recently, Qin and colleagues³ noted that the terms “immune correlate of protection” often confusingly merged three “immune correlates” definitions. First, the “correlate of risk” defined above as an immune response associated with the rate of HIV-1 infection in the vaccine group. Then, there are two levels of surrogate (i.e., a correlate that reliably predicts the level of vaccine efficacy) with increasing value: a level 1 “specific surrogate” has predictive value in the same or similar setting as the efficacy trial that yielded the surrogate, while a level 2 “generalized surrogate” (or “bridging surrogate”) has predictive value in a different setting. The search for a surrogate endpoint ultimately aims to extrapolate vaccine efficacy to new settings, with different viral strains, vaccine vectors, host backgrounds, risk groups, routes of exposure, cultural characteristics, and/or vaccine production procedures. Such generalized surrogates with cross-predictive utility are difficult to validate; meta-analyses of Phase IIb proof-of-concept and Phase III efficacy trials as well as post-licensure studies may be needed. To enable the assessment of a generalized surrogate via meta-analysis, it is critical to standardize assays and measurements of immune correlates across different studies.

The above surrogate endpoint concepts are limited to the reliable prediction of the level of vaccine efficacy, but it does not attempt to formally prove that the surrogate is the mechanistic cause of protection. We note that knowing the causal mechanism is not necessary for developing a highly efficacious vaccine; plus, efficacy trial data do not provide the requisite information to formally prove causes. To identify causation, other types of experiments that manipulate the vaccine-induced immune response in animal studies would be needed.⁴

Plotkin's⁴ definitions of correlate and surrogate differ from those of Qin et al.³ used for RV144, and to quell potential confusion we summarize the differences in definitions (Table 1). Plotkin's correlate is a measured immune response that is the mechanistic cause of protection, the actual functional mechanism by which HIV-1 infection is blocked, whereas the correlate of Qin et al. is merely a measured immune response that is statistically associated with infection rate. The former meaning seems to derive from the long tradition of field vaccinologists, while the latter meaning derives from the long tradition of statisticians to generally reserve the term correlate for a statistical association between observed variables. Coincidentally, the surrogate of Plotkin and Qin et al. means es-

entially the same thing, in each case a measured immune response that reliably predicts vaccine efficacy, without a requirement that it is the mechanistic cause of protection (a subtle difference is that the surrogate of Plotkin is *not* the mechanism of protection whereas that of Qin et al. may or may not be; Qin et al. focus on the statistical evaluation of surrogates and note that the observed trial data cannot establish or reject mechanistic causation). Given the confusion that continues to be generated by disparate meanings of the terms correlate and surrogate, there is a need for the vaccine field to develop a universal nomenclature (which is beyond the scope of this article; here we use the nomenclature of Qin et al.).

We consider that no approach is superior to the other and that they are both useful. Indeed, statistical evidence that a correlate of risk has surrogate value in the Qin et al. terminology generates the hypothesis that the correlate is the (or part of the) causal mechanism of vaccine-induced protection, i.e., a correlate in the Plotkin terminology. This hypothesis may be tested, e.g., in immune biomarker passive transfer animal challenge studies,⁴ and can guide the design of new vaccine constructs. Moreover, if a correlate of risk is identified but the study design does not permit evaluating its surrogate value (as in standard efficacy trials), the correlate can then be tested as the (or part of the) causal mechanism of vaccine protection, motivating follow-up animal challenge experiments.

Pilot Immunogenicity Studies

Before the case-control study, pilot immunogenicity studies are performed to identify the most relevant experiments among dozens of candidate assays. The pilot studies compare multiple assays to provide a rational basis for advancing specific assays; hence, pilot studies must be designed to supply sufficient comparative data. The goal is to select assays that measure relatively independent immune functions (and hence the readouts are not highly correlated with readouts from other assays) and that together cover the broadest “immunological space”; for some functions, there may be no single best assay and it may be preferable to retain several assays. The chosen assays should present low false-positive rates and relatively low noise (e.g., high reproducibility on replicate samples), while displaying large variability in vaccine-induced responses. The prototype RV144 pilot dataset comprised 100 uninfected subjects, 80 vaccine and 20 placebo recipients, with samples taken preimmunization and at peak vaccine-induced response (week 26). For vaccine recipients, week 26 responses were compared to the preimmunization readouts and to week 26 readouts in samples from placebo recipients.

TABLE 1. CORRELATE VS. SURROGATE

Correlate	Vaccine-induced immune response associated with the rate of HIV-1 infection among vaccine recipients.
Surrogate	<i>Can be identified in standard vaccine efficacy trials.</i> Correlate that reliably predicts the level of protection from HIV-1 infection. <i>Its evaluation requires subject predictors of the correlate and/or vaccination of placebo recipients.</i> <i>Provides more rigorous guidance on trial-endpoint decisions for vaccine development.</i>
Level 1 specific surrogate	Predictive value for protective efficacy in the same or similar vaccine efficacy trials.
Level 2 generalized surrogate	Predictive value for protective efficacy in different vaccine efficacy trial settings (bridging).

Definitions from Qin et al. (2007).³

An assay is selected for the case-control study based on the statistical analyses of the pilot studies, which are conducted in a standardized and uniform manner to allow the comparative interpretation of results from a variety of assays. Yet, it is undeniable that there is variation in those assays: e.g., some assays have been validated and are fairly well standardized compared to newer assays; and even within one laboratory, there can be variation between runs, between technicians, between days, and within a plate, and it is resource prohibitive to comprehensively assess all of these factors before advancing assays to the case-control study.

Ideally, there would be high variability in vaccine recipient readouts at week 26, and most of the variability would be plausibly protection relevant. In practice, it is not straightforward to understand what constitutes protection-relevant versus protection-irrelevant variability, although some measurable components of variability are clearly protection irrelevant, such as technical assay measurement error (which can be measured by evaluating the reproducibility of the assay across replicate samples) or biological variability stemming from different time intervals between the week 24 vaccination and the collection of the week 26 sample. A high fraction of protection-relevant variability makes it easier to detect an immunological measurement as a correlate, as the noise level erodes the statistical power of the correlates analysis. Therefore, if two assays testing a particular function are equally suitable on biological grounds for advancement to the case-control study, then the most reproducible assay should be given precedence. In addition, assays that require limited specimen volumes are preferred, because it is necessary to perform all the case-control assays on samples from all the RV144 vaccine recipients who became infected. Lastly, demonstrated associations with infection in previous similar efficacy trials may be informative for advancing assays to case-control.

Since there are multiple assays with partly overlapping immune characteristics, assays may be advanced to the case-control study in two tiers: a first tier comprising a limited set of priority assays and a second tier with other assays that pass a gateway criteria (this is limited by the available specimen volumes). This two-tier strategy facilitates a timely analysis and is useful for structuring the case-control statistical analysis, e.g., false-positive rates are controlled separately for tier 1 assays (below we summarize how RV144 controls for false-positive rates). While controlling for false-positive rates is important, it should not overshadow the main goal of the case-control study, which is to generate hypotheses about immune correlates of protection. False discovery rate adjustment via q -values may be preferred to the more stringent Bonferroni-type correction because the study seeks to generate hypotheses rather than a confirmation. (A q -value is the estimated expected chance that a significant result is a false positive.) In confirmatory studies (e.g., Phase III licensure trials), stringent correction (e.g., Bonferroni-adjusted p -value < 0.05) is most appropriate because a vaccine should not be licensed unless there is compelling evidence that it works. However, in hypothesis-generating studies, a 20% chance is acceptable because missing true correlates is a great concern, and requiring stringent false-positive control would increase the chance of missing correlates (i.e., reduce statistical power).

In particular, missing a true correlate of protection would have important adverse consequences, especially given that

results from the only ongoing HIV-1 vaccine efficacy trial will not be available for some years. It would deprive the vaccine field of a hypothesis that would have several applications: (1) to guide development of improved next-generation vaccine candidates, by representing the target that future vaccines (related to the RV144 regimen) would seek to match and surpass in Phase I/II vaccine trials; (2) to motivate parallel nonhuman primate studies aimed at evaluating vaccine efficacy as a function of the level of the correlate measurement; and (3) to motivate nonhuman primate studies to test the correlate as a causal mechanism of vaccine-induced protection (as mentioned above).

Generation of the hypothesis that an immune biomarker is a correlate of protection from an efficacy trial would improve the rationale for prioritizing immune biomarkers in vaccine development, given that the available rationale is limited to natural history studies, previous efficacy trials that showed no efficacy, basic science data, and animal challenge models.

Case-Control Study

The goals of the case-control study are first to assess correlates of risk of HIV-1 infection in the vaccine group, and then to test whether and how well these identified correlates can serve as a level 1 specific surrogate endpoints for HIV-1 infection.

In RV144, plasma samples were collected every 6 months from week 0 through week 182, and PBMC samples were collected at the week 0, 26, 52, and 182 visits. Accordingly, for assays performed on plasma samples, the case-control study is structured so that immune responses can be tested as potential correlates of infection in two ways: first, whether the week 26 immune response (approximate peak response) predicts infection during the next 3 years, and second, whether the most recent immune response predicts infection during the next 6 months. Only the former condition is assessed on peripheral blood mononuclear cells.

Immune responses are measured in all vaccine recipients infected after week 26 (41 individuals), and in random samples of never-infected vaccine recipients (5:1 ratio), as well as in random samples of placebo recipients infected after week 26 (20 individuals) and never-infected placebo recipients (20 individuals). Whereas the correlates analysis is based only on the vaccine group measurements, the placebo group measurements serve to verify low false-positive rates.

For the limited number of Tier 1 assays, both univariate and multivariate analyses will be performed to assess the assays individually and jointly. False-positive rates can be controlled with a procedure calibrated for multiple testing, for example, using q -value < 0.20 as a threshold. For the larger Tier 2 set, the modeling methods are tailored to handle the large number of immunological readouts.

Given the structure of the case-control study, the modeling to handle data can be complex as hundreds of immune variables (defined by factors including assay type and antigen target) potentially need to be taken into account, and several approaches can be followed. First, we can specify a small number of key variables (e.g., 4–8), either by selecting biologically important variables that performed well in the pilot studies and cover a broad immunological space or by performing dimensionality reduction on the large set of immunological variables to derive summary variables that capture

Figure 1 shows power curves for detecting a correlate of risk in RV144 for four normally distributed immunological measurements with no, low, medium, or high noise, which

Useful BIPs need to be strongly correlated with the biomarker and preferably be inexpensive such that they can be measured for most or all study subjects. The CPV design consists in immunizing a random sample of uninfected placebo recipients with the HIV-1 vaccine at the end of the trial and measuring their immune responses on the same schedule as was used for vaccine recipients. The assumptions are that the measurements are considered to be the same as what they would have been during the trial with the same distribution (i.e., age invariance is assumed). To employ CPV on RV144, as many uninfected placebo subjects (stratified by sex) as possible should be vaccinated on the RV144 schedule, with specimens collected at the key time-point week 26 in order to test the potential correlative immunologic variables on those samples. The most promising approach to evaluating a surrogate in RV144 would use both BIP and CPV; however, even if the BIP were a high-quality predictor, the analysis would have limited precision given that only 41 infected vaccine recipients are evaluable.

Given that vaccine efficacy trials are costly, demanding to volunteers, and take a long time before completion, establishing vaccine-induced immune responses as correlates and surrogates for HIV-1 infection is a central objective of placebo-controlled trials as it would provide a path for the development of future vaccines.

RV144 will have adequate statistical power to evaluate several immune responses as correlates: if associations are

sufficiently strong, it may be possible to identify one or more correlates, knowing that some correlates could act synergistically. It is, however, possible that the study of blood samples may not yield “true correlate(s)” mechanistically causative of protection but only statistical correlates that would correspond to unknown protective biomarkers in the mucosa. Mucosal biomarkers are most likely to be able to prevent the systemic dissemination of the HIV-1 infection at the mucosa, yet blood-derived variables may not necessarily overlap with the mucosal immune responses. Similarly to other vaccine efficacy trials, RV144 is underpowered to assess surrogate endpoints, which are more reliable than correlates as a basis for future vaccine trials. Nonetheless, as correlation is necessary for an immune variable to be a surrogate, identification of correlates would still be immensely valuable for the vaccine field, providing a more rational basis for the development of next-generation vaccine candidates and for immunogenicity endpoint selection; then, the surrogate value of the correlates could be evaluated in subsequent efficacy trials.

Since vaccine efficacy trials are not typically powered to identify immune correlates, it may be important to augment trial designs for improving the assessment of correlates, for example, by collecting extra data on baseline risk factors and predictors of immune responses to the vaccine, and by vaccinating placebo recipients at the end of the trial.

Acknowledgments

We thank Dr. José Esparza, Guido Ferrari, Barton Haynes, and Jerome Kim for comments on the manuscript. We thank the organizers of the Correlates of Vaccine Protection meeting, Guido Ferrari, Clive M. Gray, and Richard A. Koup, as well as Bonnie Mathieson, José Esparza, and Alan Bernstein for their contribution to the Scientific Committee of the meeting. We thank the Bill and Melinda Gates Foundation and the NIH Office for AIDS Research for funding the meeting, and the Global HIV Vaccine Enterprise and OCTAVE Project for funding and organizing a satellite workshop for Young and Early Career Investigators.

Funding to P.B.G. was provided by an NIH NIAID Grant 2 R37 AI054165-08. Funding to M.R. was provided in part by an Interagency Agreement Y1-AI-2642-12 between the U.S. Army Medical Research and Materiel Command (USAMRMC) and the National Institutes of Allergy and Infectious Diseases; and by a cooperative agreement (W81XWH-07-2-0067) between the Henry M. Jackson Foundation for the Advancement of Military Medicine, Inc., and the U.S. Department of Defense (DOD). The views and opinions expressed herein do not necessarily reflect those of the U.S. Army, the Department of Defense, or the National Institute of Health.

M.R. wrote the paper and P.B.G. designed and conducted the analysis and co-wrote the paper.

Author Disclosure Statement

No competing financial interests exist.

References

1. Prentice RL: Surrogate endpoints in clinical trials: Definition and operational criteria. *Stat Med* 1989;8(4):431–440.
2. Frangakis CE and Rubin DB: Principal stratification in causal inference. *Biometrics* 2002;58(1):21–29.
3. Qin L, Gilbert PB, *et al.*: A framework for assessing immunological correlates of protection in vaccine trials. *J Infect Dis* 2007;196(9):1304–1312.
4. Plotkin SA: Vaccines: Correlates of vaccine-induced immunity. *Clin Infect Dis* 2008;47(3):401–409.
5. Borgan, O., *et al.*: Exposure stratified case-cohort designs. *Lifetime Data Anal* 2000;6(1):39–58.
6. Gilbert, P.B., *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(5):666–677.
7. Forthal, D.N., *et al.*: Recombinant gp120 vaccine-induced antibodies inhibit clinical strains of HIV-1 in the presence of Fc receptor-bearing effector cells and correlate inversely with HIV infection rate. *J Immunol* 2007;178(10):6596–6603.
8. Follmann D: Augmented designs to assess immune response in vaccine trials. *Biometrics* 2006;62(4):1161–1169.
9. Gilbert PB, Qin L, *et al.*: Evaluating a surrogate endpoint at three levels, with application to vaccine development. *Stat Med* 2008;27(23):4758–4778.
10. Qin L, Gilbert PB, *et al.*: Assessing surrogate endpoints in vaccine trials with case-cohort sampling and the cox model. *Ann Appl Stat* 2008;2(1):386–407.
11. Gilbert PB and Hudgens MG: Evaluating candidate principal surrogate endpoints. *Biometrics* 2008;64(4):1146–1154.
12. Gilbert PB, Qin L, *et al.*: Response to Andrew Dunning’s comment on ‘evaluating a surrogate endpoint at three levels, with application to vaccine development.’ *Stat Med* 2009;28(4):716–719.
13. Wolfson J and Gilbert P: Statistical identifiability and the surrogate endpoint problem, with application to vaccine trials. *Biometrics* 2020;66(4):1153–1161.

Address correspondence to:

Morgane Rolland
U.S. Military HIV Research Program
1600 East Guide Drive
Rockville, Maryland 20850

E-mail: mrolland@hivresearch.org