What Constitutes Efficacy for a Human Immunodeficiency Virus Vaccine that Ameliorates Viremia: Issues Involving Surrogate End Points in Phase 3 Trials

Peter B. Gilbert, Victor G. DeGruttola, Michael G. Hudgens, Steven G. Self, Scott M. Hammer, and Lawrence Corey

Statistical Center for HIV/AIDS Research and Prevention, Fred Hutchinson Cancer Research Center, and Department of Microbiology, University of Washington, Seattle; Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts; Division of Infectious Diseases, Columbia University College of Physicians and Surgeons, New York, New York

Initial human immunodeficiency virus (HIV) vaccines are unlikely to prevent acquisition of HIV in all recipients. Moreover, several HIV vaccines are under evaluation that are designed to reduce viremia after acquisition of infection. Such vaccines could provide important benefits to delay HIV progression and to reduce transmission. The decision to license a vaccine on the basis of observed effects on virus load and other postinfection surrogate end points in an efficacy trial is complicated by uncertainty about whether the vaccine effects will persist and reliably predict clinical effects, and by the challenge in interpreting the data posed by treatment of some seroconverters with antiretroviral drugs. Here, we evaluate how analyses of certain surrogate end points can be used for inferring clinically significant vaccine effects and propose end points that could be evaluated in efficacy trials to support licensure. The assessment suggests that a vaccine demonstrating moderately durable effects to delay therapy and to ameliorate viremia merits consideration for licensure.

Applied widely, a vaccine highly effective in durably preventing acquisition of most human immunodeficiency virus (HIV) strains would extinguish the HIV epidemic [1–3]. However, initial vaccines designed to reduce HIV acquisition are likely to demonstrate modest levels of protective efficacy at best (30%–50%). Moreover, results from animal studies have generated enthusiasm for a whole series of candidate HIV vaccines that stimulate cell-mediated immune responses directed at controlling viral replication after acquisition of infection [4–11], and several of these vaccines are expected to enter efficacy trials in humans by 2007 [12]. When tested for efficacy, such a vaccine may reduce virus load in blood or genital secretions and delay the onset of antiretroviral therapy (ART) in vaccine recipients who become infected. Debate will ensue regarding whether such evidence is sufficient to license the vaccine for general or targeted use. The debate probably will focus on how strongly and durably the vaccine affects virus load, CD4 cell count, and other outcomes measured after infection, because certain effects on these outcomes would imply that use of the vaccine would slow disease progression in HIV-infected vaccine recipients and abate the epidemic by reducing infectiousness [13–16].

Published reports on the design of HIV vaccine efficacy trials have scarcely considered how estimated vaccine effects on virus load–based and other postinfection end points should be used to influence licensure decisions. Furthermore, treatment of those who serocon-
vert complicates measurement of vaccine effects on disease progression and secondary transmission in an efficacy trial; to our knowledge, no proposals have been offered to account for treatment of seroconverters in the design or analysis. In the present article, we address assessment of prophylactic vaccine effects on disease and transmission on the basis of the data expected from an efficacy trial and the fact that some HIV-infected participants will initiate treatment. We propose use of specific surrogate end points, with emphasis on a composite end point defined as either loss of virologic control or treatment initiation; this end point is clinically interpretable and accommodates treatment of HIV-infected participants, which is an important ethical consideration. Our investigation of the potential benefits of the surrogate end points centers on a numerical study of a hypothetical HIV vaccine efficacy trial design with characteristics similar to those of the 2 ongoing efficacy trials [17, 18]. From this numerical study, we identify vaccine effects on the proposed end points that, if observed in an actual trial, would be expected to imply clinically significant vaccine effects on disease and transmission and, therefore, would form the basis for licensing an HIV vaccine.

This article is organized into 3 sections. In the first section, we discuss objectives of HIV vaccine efficacy trials and end points used to address the objectives, with a focus on appropriately accounting for treatment initiation. With assistance from the numerical study, in the second section, we address the question of whether and how surrogate end points can be used to identify robust and clinically significant vaccine effects on disease and transmission. In the third section, we draw conclusions and discuss requisite long-term follow-up studies of HIV-infected participants and phase 4 studies that would follow efficacy trials that led to a licensed vaccine. This article focuses on measuring effects of prophylactic vaccines (administered to HIV-uninfected individuals) but does not consider therapeutic vaccines (administered to HIV-infected individuals).

**SURROGATE END POINTS IN PROPHYLACTIC HIV VACCINE EFFICACY TRIALS**

Forthcoming efficacy trials are expected to have the following design characteristics. Several thousand HIV-uninfected volunteers are randomized to receive vaccine or a placebo (control) and are monitored semiannually for HIV infection during a 3–4-year period. Subjects who seroconvert are monitored for virologic, immunologic, and clinical outcomes periodically (e.g., monthly for the first 6 months after detection of HIV and semiannually thereafter for ≧5 years).

**Clinical vaccine efficacy parameters.** In a classically designed phase 3 trial of a vaccine for a general infectious disease, the primary end point is onset of clinically significant disease [8]. The primary analysis of the trial estimates a parameter, VEₜ, defined as the percentage of reduction in the risk of the disease end point for vaccine versus placebo recipients. For a classically designed HIV vaccine trial, HIV acquisition may serve as the primary disease end point [19]. When feasible, the trial also should allow for comparisons of the rates of HIV-related morbidity and mortality events between all vaccine and placebo recipients in the study [20], as well as for comparisons of the rates of these disease events and of secondary transmission events between the subgroups of vaccine and placebo recipients who become HIV infected during the trial. The comparisons within the subgroups of HIV-infected participants would allow for assessments of clinical parameters VEₚ and VEₜ that measure vaccine efficacy to reduce progression and infectiousness: VEₚ is the percentage of reduction in the risk of AIDS or death by a time point well beyond HIV infection and VEₜ is the percentage of reduction in the risk of transmitting HIV to others [21] (table 1). Because progression of HIV often takes several years, a definitive answer on VEₜ probably will not be available until HIV-infected participants have been monitored for at least 5–7 years, even if ART is seldom used [22, 23]. Direct evaluation of VEₚ within a classical trial design requires the monitoring of sex partners of HIV-infected participants for infection, and observation of many secondary transmission events that occur, despite ongoing HIV testing and risk-reduction counseling [21, 24]. Because rates of secondary HIV transmission are usually low (e.g., 22% in the Rakai study [14, 15]), this latter task is challenging and has been reviewed elsewhere by others [25].

**Surrogate vaccine efficacy parameters.** Historically, in the absence of a known correlate of protective immunity, the standard for vaccine licensure required by regulatory agencies has been direct demonstration of clinical benefit to reduce morbidity/mortality [26]. For an HIV vaccine without definitive VEₚ, this standard could be met by showing substantial VEₚ and VEₜ combined with demonstration that susceptibility to infection is not increased by vaccination. Because of the difficulties in demonstrating effects on VEₚ and VEₜ directly, it may also be important to analyze surrogate end points of progression and transmission that can be evaluated more quickly and easily. Trials showing positive vaccine effects on surrogate end points may motivate the conduct of studies with more complicated or long-term designs that provide direct assessments of VEₚ and VEₜ; such studies are important for understanding the surrogate vaccine effects. Furthermore, if vaccine effects on surrogate end points reliably predict VEₚ and VEₜ, then the surrogates might be relied upon as primary end points for supporting licensure decisions. Clinical trials of ARTs for HIV-infected individuals represent an example in which surrogate end points (virus load–based) were used as the primary basis for licensure decisions; combinations of drugs licensed after such trials led to large population reductions in rates of
AIDS and death [27–32] (analogous to VE<sub>P</sub>). Some of the ART trials demonstrated treatment effects on both surrogate and clinical end points [30, 33], which provided the basis for investigating the extent to which surrogate vaccine effects reliably predicted clinical vaccine effects. In contrast to the apparent success of ART trials, in several trials in a variety of disease areas, the use of surrogate end points misled about treatment effects on clinical end points [34–40]. Most of these misleading trials evaluated therapies for chronic diseases, for which the etiology was poorly understood. In contrast, the etiologic agent of HIV disease is known, and the extensive, albeit imperfect, knowledge of HIV pathogenesis and transmission may help in defining reliable surrogate end points. Therefore, surrogate end points hold promise for use in HIV vaccine trials, provided that they are carefully chosen and evaluated and that the relationship among vaccine, surrogate, and clinical end points is fully investigated.

Here, we consider 2 types of potential postinfection surrogate end points in efficacy trials: (1) end points measurable in all HIV-infected subjects before treatment is initiated and (2) “late” end points, measured several months or years after infection that may be affected by use of treatment. An example of an early surrogate end point is initial pretreatment virus load. The potential usefulness of this end point is based on the observation that it is highly prognostic for AIDS and death [22, 23, 38, 41–43] and for heterosexual transmission [14, 15]. For some seroconverters, the virus load at the time of the first positive HIV test may be highly variable, but alternatively to defining VE<sub>P</sub> based on infectious contacts occurring at T months, VE<sub>P</sub> can be defined based on infectious contacts occurring during a time period T, to T months (e.g., 1–8 months after infection diagnosis).

### Table 1. Definitions of clinical and surrogate vaccine efficacy parameters.

<table>
<thead>
<tr>
<th>Vaccine efficacy parameters</th>
<th>Analyzed cohort</th>
<th>Time period for measuring end points</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Clinical</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VE&lt;sub&gt;P&lt;/sub&gt;(&lt;sub&gt;T&lt;/sub&gt;): vaccine efficacy susceptibility</td>
<td>Randomized cohort</td>
<td>Between entry and T months after entry (e.g., T = 36 months)</td>
</tr>
<tr>
<td>Percentage of reduction in the risk of acquiring HIV infection by &lt;sub&gt;T&lt;/sub&gt; months for vaccine vs. placebo recipients</td>
<td>Infected subcohort</td>
<td>Between infection diagnosis and T months after infection diagnosis (e.g., T = 84 months)</td>
</tr>
<tr>
<td>VE&lt;sub&gt;P&lt;/sub&gt;(&lt;sub&gt;T&lt;/sub&gt;): vaccine efficacy disease progression</td>
<td>Infected subcohort</td>
<td>At T months after infection diagnosis (e.g., T = 3 months)</td>
</tr>
<tr>
<td>Percentage of reduction in the cumulative risk of progressing to AIDS or death by &lt;sub&gt;T&lt;/sub&gt; months for vaccine vs. placebo recipients</td>
<td>Infected subcohort</td>
<td>Between entry and T months after entry (e.g., T = 36 months)</td>
</tr>
<tr>
<td>VE&lt;sub&gt;P&lt;/sub&gt;(&lt;sub&gt;T&lt;/sub&gt;): vaccine efficacy infectiousness</td>
<td>Infected subcohort</td>
<td>Between infection diagnosis and T months after infection diagnosis (e.g., T = 18 months)</td>
</tr>
<tr>
<td>Percentage of reduction in the per-contact probability of transmitting HIV to others at &lt;sub&gt;T&lt;/sub&gt; months for vaccine vs. placebo recipients</td>
<td>Infected subcohort</td>
<td>Between infection diagnosis and T months after infection diagnosis (e.g., T = 18 months)</td>
</tr>
</tbody>
</table>

| Surrogate                  |                |                                     |
| ΔVL<sub>P</sub>(<sub>T</sub>): vaccine efficacy pretreatment virus load | Infected subcohort | Between infection diagnosis and T months after infection diagnosis (e.g., ΔVL<sub>3</sub> = vaccine efficacy on initial pretreatment virus load) |
| Mean reduction in virus load measured before treatment at <sub>T</sub> months for vaccine vs. placebo recipients | Infected subcohort | Between infection diagnosis and T months after infection diagnosis (e.g., T = 18 months) |
| ΔCD4(3)<sub>T</sub>: vaccine efficacy pretreatment CD4 cell count | Infected subcohort | Between infection diagnosis and T months after infection diagnosis (e.g., T = 18 months) |
| Mean reduction in CD4 cell count measured before treatment at <sub>T</sub> months for vaccine vs. placebo recipients | Randomized cohort | Between entry and T months after entry (e.g., T = 36 months) |
| VEVL<sub>P</sub>(<sub>T</sub>;<sub>X</sub>): vaccine efficacy virologic failure or treatment initiation susceptibility | Infected subcohort | Between infection diagnosis and T months after infection diagnosis (e.g., T = 18 months) |
| Percentage of reduction in the cumulative risk of virologic failure (virus load &gt;X e.g., X &gt; 55,000 copies/mL) or treatment initiation (whichever occurs first) by <sub>T</sub> months for vaccine vs. placebo recipients | Randomized cohort | Between entry and T months after entry (e.g., T = 36 months) |
| VECD4<sub>P</sub>(<sub>T</sub>;<sub>X</sub>): vaccine efficacy CD4 cell count failure or treatment initiation susceptibility | Infected subcohort | Between infection diagnosis and T months after infection diagnosis (e.g., T = 18 months) |
| Percentage of reduction in the cumulative risk of CD4 failure (CD4 cell count &lt;X e.g., X &lt; 350 copies/mm<sup>3</sup>) or treatment initiation (whichever occurs first) by <sub>T</sub> months for vaccine vs. placebo recipients | Randomized cohort | Between entry and T months after entry (e.g., T = 36 months) |

* Alternatively to defining VE<sub>P</sub>(<sub>T</sub>) based on infectious contacts occurring at T months, VE<sub>P</sub>(<sub>T</sub>) can be defined based on infectious contacts occurring during a time period T, to T months (e.g., 1–8 months after infection diagnosis).

* Alternatively to defining ΔVL<sub>P</sub>(<sub>T</sub>) or ΔCD4<sub>P</sub>(<sub>T</sub>) based on a single biomarker measurement at T months, these parameters can be defined based on the average of multiple measurements of the biomarker during a time period T, to T months.
because HIV was recently acquired [20]. To minimize variability of the end point, it can be defined using only pretreatment measurements made at least 1 month after the initial positive test result for HIV. A surrogate vaccine efficacy parameter, ΔVL(T), can be defined as the mean pretreatment virus load at T months after HIV diagnosis (e.g., T = 3 months) in HIV-infected placebo recipients minus that in HIV-infected vaccine recipients (table 1). Generally the interpretation of early end points will not be complicated by treatment, but these end points provide little direct information about the durability of effects and, therefore, will have a more tenuous connection to later clinical events in the course of infection.

Initial vaccine efficacy on surrogate end points may wane because of emergent HIV vaccine resistance mutations [44, 45]. For example, vaccine control of viremia may be lost because of viral escape from cytotoxic T lymphocyte (CTL) responses [46–49]; HIV CTL escape mutants have been linked with a higher plasma virus load and more-rapid disease progression [48] and have been documented to cause simian-HIV (SHIV) vaccine failure [50]. Moreover, such mutants are fully virulent and can be transmitted [49]. Therefore, to ensure some durability of vaccine effects, “late” surrogate end points must be evaluated to support licensure decisions. Late end points based on CD4 cell count and virus load are potentially useful [14, 15, 22, 33, 38, 41–43, 51–55], but the ethically mandated provision of ART when HIV-infected subjects reach treatment criteria poses a difficult challenge to their definition and analysis.

Accommodating ART initiation in defining a late surrogate end point. If no HIV-infected participants initiate ART, then the choice of late surrogate end points is simplified. End points commonly used in ART trials could be used, such as trajectories of virus load and CD4 cell counts or so-called virologic failure, which is the time from the diagnosis of HIV infection until virus levels increase above a failure threshold (X copies/mL) [56]. However, even in this simple situation, the reliability of such surrogates is uncertain, because surrogacy ultimately depends on the mechanism of action of the intervention under study, and the mechanisms of vaccine effects are very different than those of ART. Nonetheless, a substantial and durable suppression of virus may serve as the basis of a compelling case for licensure, but the questions that remain are how large and how durable.

If some HIV-infected participants begin potent ART, the resulting suppression of virus load to undetectable levels [33, 51, 57, 58] may mask a vaccine effect in the treated groups. Consequently, an analysis of virus load including all HIV-infected subjects, regardless of ART use, would be difficult to interpret. To illustrate the problem, suppose the vaccine effectively controls virus loads of most HIV-infected vaccine recipients to 10,000–50,000 copies/mL, but, in HIV-infected placebo recipients, the virus load frequently exceeds 55,000 copies/mL, the threshold at which treatment is recommended [59]. Because more placebo than vaccine recipients will initiate ART, the virus load will tend to be lower in the placebo group at some time points. An analysis that includes subjects receiving ART would make the vaccine appear inferior to placebo for durably suppressing viremia, an incorrect inference about efficacy of the vaccine to be used in a population not receiving ART. A second approach would study the time until virologic failure, with treated individuals censored from the analysis at treatment onset. This approach is complicated because the chance of initiating ART probably depends on the risk for virologic failure [60], thereby invalidating standard analytic methods that assume independent censoring (e.g., Kaplan-Meier and Cox regression methods).

We propose use of a composite virus load end point “C” defined as the first event of virologic failure (>X copies/mL) or treatment initiation; this end point can be assessed by use of standard methods. The composite end point is directly tied to clinical events, because virologic failure places a subject at increased risk for progression and transmission, and initiating ART exposes a subject to drug toxicities, drug resistance, and the loss of future therapy options [59–61]. Furthermore, absence of the composite end point is clearly beneficial for a person, indicating viremic control without requiring therapy. The composite end point measures the magnitude of viremic control through the choice of failure threshold X, with a significant vaccine effect for lower thresholds indicating greater control (e.g., virologic control <X = 1500 copies/mL could practically eliminate secondary transmission [14, 15]). In addition, the end point captures durability of the vaccine effect by including failure events during a sufficiently long period of 7 months after infection diagnosis; ultimately, several years may be required to ensure the end point’s clinical relevance. The composite end point can be analyzed to assess a surrogate vaccine efficacy parameter VEVL(T,X), defined as the percentage of reduction in the risk of failure (either virologic failure >X copies/mL or ART initiation) for vaccine versus placebo recipients by T months after infection detection (table 1).

At a given analysis time, the duration of follow-up of HIV-infected participants dictates the maximum monitoring period T that can be used for capturing postinfection end points; therefore, the first analysis of these end points should be timed to provide enough information on the durability to potentially support a licensure decision. We suggest that follow-up of 18 months after infection diagnosis may be a reasonable minimum duration, because natural history data suggest that virus load at 15–18 months after seroconversion is more predictive of progression than virus load at 1–5 months after seroconversion [23], and relatively few persons are expected to qualify for ART within 18 months [59, 62]. Furthermore, experience with antiretroviral drugs suggests that even modest decreases in virus load during
the first 6 months after initiation of therapy in established infection are associated with lower rates of progression over 1–3 years of follow-up [53]. If this phenomenon also occurs for vaccine effects, then a vaccine-induced suppression of viremia over 18 months would be expected to translate into clinical benefit. Accordingly, we focused on the assessment of $\text{VEVL}_c(T;X)$, with $T = 18$ months, while emphasizing that all seroconverters should be monitored for at least 5 years and that $\text{VEVL}_c(T;X)$ be analyzed with $T$ as large as feasible at each analysis time [e.g., update analyses of $\text{VEVL}_c(36,X)$ and $\text{VEVL}_c(60,X)$].

Figure 1 illustrates the analysis of $\text{VEVL}_c(18,X)$ for a data set simulated under trial parameters described in the following section (Analysis of Surrogate End Points for Licensure Decisions of Candidate HIV Vaccines). To convey the magnitude of effect, $\text{VEVL}_c(18;X)$ was estimated for each of a series of failure thresholds $X$, with 95% confidence intervals (CIs). Note that for all threshold values of $X > 5000$ copies/mL, the lower limit of the 95% CI for $\text{VEVL}_c(18;X)$ was >40%. We argue in the following section that observation of such an effect in an actual trial would make it reasonable to infer clinically significant vaccine efficacy to delay disease and to reduce transmission.

The choice of threshold $X$ and the interpretation of the composite end-point analysis depend on the treatment initiation guidelines that are used in the trial and adherence to those guidelines. For example, if the guidelines specify initiating ART when virus load is >55,000 copies/mL, then it is sensible to estimate $\text{VEVL}_c(18;X)$ for $X$ values ranging up to 55,000 copies/mL; if the guidelines are followed, each estimate has clear interpretation as the effect of vaccination on the virologic failure rate without confounding by treatment. Alternatively, the treatment initiation guidelines might be based on CD4 cell count or on both CD4 cell count and virus load; for example the current US guidelines recommend therapy when the CD4 cell count is <350 cells/mm$^3$ or virus load is >55,000 copies/mL [59]. Such a guideline is based on loss of either immune competence or virologic control. This guideline complicates the interpretation of the composite end-point analysis, because some subjects may initiate ART before virologic failure. Assessments of vaccine effects on the specific event types of virologic failure, CD4 cell count below the treatment threshold of 350 cells/mm$^3$, and ART initiation outside the guidelines that compose the composite end point help interpret the analysis.

Achieving high rates of adherence to the treatment initiation guidelines makes the composite end-point analysis easier to interpret. Adherence can be promoted within the trial design by providing free ART to all participants who reach treatment cri-

![Figure 1. Analysis of $\text{VEVL}_c(18;X)$ (vaccine efficacy virologic failure $> X$ or treatment initiation within 18 months) for a range of virologic failure thresholds $X$. For a vaccine trial data set simulated under design parameters given in figure 2, the bold line indicates the estimated $\text{VEVL}_c(18;X)$, plotted vs. the virologic-failure threshold $X$ used in the end-point definition. For a given threshold $X$, $\text{VEVL}_c(18;X)$ is estimated by $1 - \text{vaccine vs. placebo ratio of the percentage of human immunodeficiency virus–infected participants who experienced the composite end point by 18 months after infection detection (each percentage is estimated by the Kaplan-Meier method). The shaded region represents 95% confidence intervals for $\text{VEVL}_c(18;X)$.)
taria; HIV Vaccine Trials Network studies employ this policy. If a moderate fraction of seroconverters initiate ART before meeting treatment criteria (e.g., <50%), then the composite end-point analysis is still expected to be useful. However, if most seroconverters initiate ART before qualification, the composite end-point analysis contributes little independent information beyond the analysis of treatment initiation, regardless of reason. For this case, the assessments of initial pre-ART virus load and CD4 measured at least 18 months after infection diagnosis may be more informative than the composite end-point assessment. In general, a high rate of early treatment initiation, regardless of clinical and biomarker disease indicators, would reduce the power and complicate the interpretation of any analytic approach based on pre-treatment surrogate end points.

Even with perfect adherence to the guidelines, the rate and timing of treatment initiation affect the analysis of pre-ART surrogate end points, with a lower rate and a later onset of therapy allowing for more powerful assessments. For a trial conducted according to current US guidelines, 10%–30% of seroconverters might be expected to initiate ART within 18 months [59], whereas, for a trial in a resource-limited setting conducted according to current World Health Organization guidelines, a lower rate is expected [62]. The US projection includes HIV-infected participants who initiate ART very early because of detection of infection in the acute phase [59]. Although such HIV-infected participants contribute little information about pre-ART vaccine effects, the resulting impact on the analysis is likely minor, because only 5%–10% of HIV-infected participants are expected to be diagnosed in the acute window, given a typical semiannual HIV-testing schedule [63]. Furthermore, the trial could be powered to analyze the subgroup of subjects diagnosed after the acute stage. Because the rate and timing of treatment initiation may be unpredictable at the onset of a trial [64], an advantage of the composite end point is that its analysis is statistically valid for any rate and timing of initiating ART. The analysis of VE\(V_{L}(18;X)\) will be most interpretable for measuring the vaccine effect on virus load if \(X\) is set low (e.g., \(X < 5000\) copies/mL), because, in this case, a relatively large fraction of the composite end points will be caused by virologic failure.

**Other late surrogate end points.** The inference on VE\(V_{L}(18;X)\) applies to a population with treatment initiation guidelines and patterns similar to those observed in the trial. To complement VE\(V_{L}(18;X)\), we also considered the surrogate vaccine-efficacy parameter \(\Delta V_{L}(T)\) for \(T\) a late time-point such as 18 months, defined as the mean difference in virus load in placebo versus vaccine recipients measured \(T\) months after infection diagnosis, for a hypothetical setting in which no seroconverters began treatment by \(T\) months (table 1). The inference on \(\Delta V_{L}(T)\) applies to a population who would not use ART during the first \(T\) months after diagnosis of infection. \(\Delta V_{L}(T)\) for a series of visit times \(T\) can be assessed by use of statistical methods that censor virus-load trajectories at the time of ART initiation and model the dependency of the chance of initiating ART on virus load and other factors. Two modeling approaches can be used for estimating \(\Delta V_{L}(T)\). If there is low- to-moderate adherence to the treatment guidelines, then the method described by Rotnitzky et al. [65] can be applied, which models the probability of observing a pre-ART virus load at each visit time with a logistic regression model, and models the mean pre-ART virus load with a regression model (e.g., a linear model). This approach extends the commonly used method of generalized estimating equations [66] by weighting observations by estimated probabilities of observing pre-ART virus loads. If there is good adherence to treatment guidelines, this method, however, is inapplicable, because the weights are nearly zero, which makes the estimates of \(\Delta V_{L}(T)\) unstable. In this case, analyses can be based on standard linear mixed models that correct for dependent censoring by incorporating predictors of treatment initiation [67]. These models require more modeling assumptions than the method described by Rotnitzky et al. [65].

Because of the prognostic value of CD4 cell count, we also considered parallel surrogate end points based on CD4 cell counts, such as pre-ART CD4 cell count at 18 months and the time until CD4 cell counts decreased to <350 cells/mm\(^3\), which could be assessed regardless of ART use or as a composite end point with treatment initiation (table 1). CD4-based end points may be preferable to virologic-based end points as late surrogate end points for disease progression; this conclusion is based on studies that showed, near the time of AIDS diagnosis, CD4 cell counts are better predictors of AIDS than virus load [23, 38, 68]. On the other hand, the same studies supported early virus load as more predictive of progression than early CD4 cell count, and the Rakai study [14, 15] and perinatal transmission trials [69] suggested that virologic-based end points may be superior to CD4-based end points for marking infectiousness throughout the progression period [14, 15].

Other biomarker end points will be studied in efficacy trials including serial measurements of genotypic/phenotypic properties of infecting viruses [45, 70] and of the magnitude and breadth of immune responses to certain HIV epitopes evaluated in relation to host factors such as HLA type. Both HIV-specific CD4 and CD8\(^+\) T cell responses will be measured. In the analyses of the main surrogate end points, it may be important to control for these factors as well as for other potential host and viral predictors of the surrogate and clinical end points, such as sex [71], behavioral variables [14], and chemokine receptors used for HIV entry into cells (e.g., CCR5 or CXCR4) [72, 73]. The time until treatment initiation irrespective of biomarker values also should be studied; analysis of these times assesses vaccine-induced reductions in therapy use, with attending sav-
A "false-positive" decision to license and distribute a vaccine that later is proven ineffective could have high manufacturing and political costs and could impede other HIV vaccine trials.

To balance the pros and cons, our approach identifies the size and duration of the observed effects on the surrogate end points required to reasonably infer that the risk of a false-positive licensure decision is low. To illustrate points, we considered a specific hypothetical trial design that reflects the design of the 2 ongoing efficacy trials [17, 18] (figure 2). Accrual of participants takes place over 12 months. Participants are evenly randomized to receive vaccine or placebo and are monitored for HIV infection for 36 months after enrollment. The trial is designed so that 125 placebo recipients are expected to acquire HIV infection. A single interim analysis is planned to occur halfway through the trial (i.e., 3 months after 63 placebo recipients have acquired HIV infection), and a final analysis of VE\textsubscript{S} is planned to occur 3 months after the last enrolled participant has been monitored for 36 months (expected 51 months into the trial). This design has 80% power to detect VE\textsubscript{S} >30%, if the true VE\textsubscript{S} is \geq 60% [19]. A final analysis of the postinfection surrogate end points is planned to occur 3 months after all seroconverters have been followed for at least 18 months after infection diagnosis.

Figure 2 shows "decision matrices" for the interim and 2 final analyses, based on joint analysis of VE\textsubscript{S} and the surrogate parameters for VE\textsubscript{P} and VE\textsubscript{E}. Each matrix divides assessment of vaccine effects into effects on the infection end point (VE\textsubscript{E}) and effects on the postinfection end points (surrogates for VE\textsubscript{P}/VE\textsubscript{S}). For each component, at each analysis time, the information from the trial suggest 1 of 3 decisions: "+" indicates clinically significant efficacy, which may support licensure; "−" indicates that the vaccine is harmful (e.g., a higher infection rate or a higher virus load in vaccine recipients), or "−" indicates an inconclusive result. For VE\textsubscript{E}, the "+", "−", and "×" decisions may be determined by whether the lower 95% CI limit for VE\textsubscript{E} is >30%, the upper 95% CI limit is <0%, or otherwise, respectively [19]. Here, we developed decision criteria for the postinfection component VE\textsubscript{P}/VE\textsubscript{E} by identifying how large the estimated surrogate parameters \(\Delta VL(3)\) and \(VEVL_c(18;X)\) must be to reliably infer clinically significant magnitudes and durations of VE\textsubscript{E} and VE\textsubscript{P}. It is also important to evaluate the corresponding CD4-based parameters; however, for brevity, we focused on the virologic measures. We first noted that very limited information about the durability of postinfection vaccine effects would be available at the time of the interim analysis (figure 2) and therefore suggested that only VE\textsubscript{E} be used in formal guidelines for stopping the trial early for evidence of efficacy (as was done for the ongoing trials [17, 18]).

**Numerical study.** We used a numerical study of the trial described in figure 2 to assess how well VE\textsubscript{E}(3) and VE\textsubscript{P}(84) can be predicted from an estimate of \(\Delta VL(3)\), as well as how well VE\textsubscript{E}(84) can be predicted from an estimate of \(VEVL_c(18;X)\). The prediction of VE\textsubscript{E}(3) from \(\Delta VL(3)\) is based on the Rakai study [14], which estimated that the risk of heterosexual HIV transmission is reduced 2.45-fold per 1-log\textsubscript{10} lower plasma virus load of the exposing partner. Specifically, VE\textsubscript{E}(3) is predicted as follows: Predicted VE\textsubscript{E}(3) = \left[ 1 - \left( 1 - 0.45^{\Delta VL(3)} \right) \right] \times 100\%.

The prediction of VE\textsubscript{P}(84) from \(\Delta VL(3)\) is based on a meta-analysis of 3 natural history studies of HIV-infected individuals who did not receive ART; these studies estimated a 2.39-fold reduction in the risk of progression to AIDS or death per 1-log\textsubscript{10} lower initial plasma virus load (with adjustment for baseline CD4 cell count) [43]. This result leads to the following equation: Predicted VE\textsubscript{P}(84) = \left[ \frac{1 - \left[ 1 - \left( 1 - PR_h \right)^R \right]}{PR_{c}} \right] \times 100\%,

where \( R = (1/2.39)^{\Delta VL(3)} \) and \( PR_h \) is the progression rate in the placebo group by 84 months. The prediction of VE\textsubscript{P}(84) from VE\textsubscript{E}(3) is based on the following equation: Predicted...
Figure 2. Schema and decision matrices for a prophylactic human immunodeficiency virus (HIV) vaccine efficacy trial. The top half of the figure depicts an example of an efficacy trial design. Accrual is uniform over 12 months, the interim analysis is triggered by the 63rd placebo infection and is expected at 28 months, the final analysis for VEₜ (vaccine efficacy susceptibility) is triggered when all participants have been followed for 36 months and is expected at 51 months, and the final analysis for surrogates of VEp/VEI (vaccine efficacy disease progression/vaccine efficacy infectiousness) is 18 months after the final analysis for VEₜ. An estimated VEₜ of 40% is assumed for calculating the number of vaccinated persons infected or with 18 months of follow-up after infection diagnosis at the analysis times. In the bottom half of the figure, “+,” “○,” and “−” indicate definitive evidence of efficacy, inconclusive evidence, and definitive evidence of harm, respectively; “+” or “−” at the interim analysis suggest the trial should be stopped early. Criteria for the “○” row, “+” column decisions in the left and middle matrices (shaded cells) are the focus of this article: What constitutes definitive evidence on the surrogates of VEp/VEI, given inconclusive evidence on VEₜ? For the shaded cell in the left matrix, we propose that no amount of evidence would be sufficient, because too little information on persistence of vaccine effects is available. For the shaded cell in the middle matrix, specific vaccine effects that would constitute definitive evidence are proposed in Analysis of Surrogate End Points for Licensure Decisions of Candidate HIV Vaccines.

VEₜ(84) = VEVLₜ(18;X). Natural history data support this prediction for X between 5000 and 10,000 copies/mL [59].

Data from the efficacy trial provide estimates of ΔVL(3), VEVLₜ(18;X) and PRₓₜ, which can be substituted into the above equations to estimate Predicted VEₜ(3) and Predicted VEp(84). The basic prediction equations are modified to account for uncertainty sources (1)–(4) listed above. We describe heuristically how this is done, with mathematical details provided in the Appendix. For uncertainty source (1), published variance estimates of the surrogate parameter estimates determine the variance estimates of the predictions. The effect of uncertainty source (2) is accounted for by conservatively assuming that the actual vaccine effect on the surrogate end point is less than the estimated effect, by an amount that is an upper bound for the plausible impact of selection bias. To accommodate the possibility that unreliability of the surrogate end points (uncertainty source 3) leads to overly optimistic predictions of VEₜ and VEp, a range of values for a bias correction factor f > 0 was specified, and the surrogate parameter estimates in the prediction equations were replaced by reduced estimates governed by f, with larger f representing more conservatism. The fraction f is the percentage of reduction of the surrogate parameter estimate, after adjustment for possible selection bias, and is determined by the following: (A) specific assumptions on the reliability of the surrogate vaccine effect to correctly predict VEₜ or VEp ("poor," "moderate," "good," or "perfect") and (B) the extent to which unreliability of the surrogate end point leads to overprediction of VEₜ or VEp ("substantial" or "extreme" overprediction). The Appendix defines the terms in quotations and lists the 8 values of f determined by the 8 combinations of (A) and (B) defined above. Adjustments made to accommodate uncertainty sources (2) and (3) cause the predicted VEₜ(3) and VEp(84) to be smaller than predicted from
the unadjusted surrogate parameter estimates obtained from the trial. Uncertainty source (4) is addressed by only using data available at the time of the specified analysis.

We summarize the main results from the numerical study. First, at the final analysis of VE_s, figure 3 shows the predicted VE_s(3) as a function of the lower 95% CI limit for ΔVL(3). Each panel represents a different combination of (A) and (B) defined above. Note that as the predictive reliability of the surrogate endpoint increases, a smaller effect on virus load is needed to infer a substantial VE_s(3). Similarly, a smaller effect is needed under an assumption of less predictive bias. For the same combinations of (A) and (B), figure 4 shows the predicted VE_s(84) from the lower 95% CI limit for VEVL_c(18;X).

For drawing specific guidelines, we focused on the predictions illustrated in the plots in the first row and second column of figures 3 and 4, which assume a moderately reliable predictive surrogate and allow for substantial overprediction. This scenario includes a reasonable degree of conservatism ($f = 0.25$) without being so conservative as to allow for a high risk of missing an effective vaccine. For example, if the estimate of ΔVL(3) (the observed mean difference in virus load) is $1.5 \log_{10}$, then, in this scenario the reduced estimate of $0.88 \log_{10}$ is used in the predictive equation, and, if the estimate of VEVL_c(18;X) is 65%, then the reduced estimate of 48% is used. From this panel in figure 3, a lower 95% CI limit for ΔVL(3) $\geq 1.3 \log_{10}$ reliably predicts VE_s(3) $\geq 40\%$, based on the lower 95% prediction limit for VE_s(3). From this panel in figure 4, a lower 95% CI limit for VEVL_c(18;X) $\geq 42\%$ reliably predicts VE_s(84) $\geq 40\%$. In addition, a lower 95% CI limit for ΔVL(3) $\geq 1.7 \log_{10}$ reliably predicts VE_s(84) $\geq 40\%$ (data not shown). The threshold guidelines are stated in terms of lower 95% CI limits so that the results are applied as generally as possible to trials of different sample sizes, virus load distributions, and composite end-point rates.

**Figure 3.** Predicted VE_s(3) from the estimated ΔVL(3) (vaccine efficacy initial pretreatment virus load) at the final analysis of VE_s. Solid lines indicate VE_s(3) (vaccine efficacy infectiousness) predicted from the lower limit (LL) of the 95% confidence interval (CI) for ΔVL(3), the vaccine efficacy to decrease initial pretreatment virus load; dotted lines are 95% prediction intervals for VE_s(3). From left-to-right, the surrogate end point is assumed to explain 25%, 50%, 75%, or 100% of the vaccine effect on infectiousness—i.e., the virus load end point is “poor,” “moderate,” “good,” or “perfect,” respectively, for predicting VE_s(3). If the surrogate end point is imperfect, then the surrogate vaccine effect may overpredict VE_s(3). The predictions correct for an assumed bias to overpredict; this bias is assumed to be substantial (top panels) or extreme (bottom panels). For each panel, the value of the correction factor $f = b(1 - p)$ is listed.
Figure 4. Predicted VEₐ from the estimated VEVL(18;X) [vaccine efficacy virologic failure >X or treatment initiation within 18 months] at the final analysis of VEₑ. Solid lines indicate VEₑ(84) [vaccine efficacy disease progression within 84 months] predicted from the lower limit (LL) of the 95% confidence interval (CI) for VEVL(18;X), the vaccine efficacy to prevent the composite virus load end point by 18 months after infection diagnosis for a particular virologic failure threshold X. Dotted lines are 95% prediction intervals for VEₑ(84). Half the human immunodeficiency virus–infected placebo recipients are assumed to experience the composite end point by 18 months. From left-to-right, the surrogate end point is assumed to explain 25%, 50%, 75%, or 100% of the vaccine effect on disease progression—i.e., the composite end point is “poor,” “moderate,” “good,” or “perfect,” respectively, for predicting VEₑ(84). If the surrogate effect is imperfect, then the surrogate vaccine effect may over-predict VEₑ(84). The predictions correct for an assumed bias to over-predict; this bias is assumed to be substantial or extreme. For each panel, the value of the correction factor \( f = b(1 - p) \) is listed.

On the basis of simulation experiments demonstrating the beneficial impact of using a vaccine with moderate VEₑ and VEₑ [76], these results suggest that the composite virus load end point could potentially be used as a coprimary end point with HIV infection, with specified effect size enough to infer a clinically significant level of efficacy. The composite end point is preferable to the initial pretreatment virus load end point, because it measures the durability of the vaccine effect. Consequently, at the final analysis of VEₑ, a vaccine showing inconclusive VEₑ and strong virologic control without treatment for at least 18 months may warrant licensure. Specifically, we propose that the “+” decision for the postinfection component of the middle matrix in figure 2 (shaded cell) may be defined by the lower 95% CI limit for VEVL(18;X) >40%, plus consistent effects on CD4 cell counts. If this positive result occurs, the final analysis of the surrogates of VEₑ/VEₑ is important for confirming that the vaccine effect persists to 36 months.

What consistency of evidence across the postinfection analyses constitutes robust support for vaccine efficacy? We propose requiring positive vaccine effects on both the virus load composite end point and the CD4 composite end point that are (1) substantial in magnitude and durability and (2) evident in analyses of all enrolled subjects and in analyses of the subcohort of seroconverters. The analyses in all subjects study the time from randomization until the composite end point, and are important because they provide unbiased inferences (by virtue of being intent-to-treat) and they approximate a classical assessment of vaccine efficacy to prevent clinically significant disease [8]. The subcohort analyses are important because vaccine effects on HIV pathogenesis are most clearly measured in HIV-
Design of HIV Vaccine Trials • JID 2003:188 (15 July) • 189

infectcd persons. In addition, the HIV-infected subcohort can be studied intensively for several years, whereas the entire cohort is too large to be studied feasibly long-term. Criteria (1) and (2) could be met if the following 4 results all occur:

1. The lower 95% CI limit for VEVL\(_C\)\((T;X)\) > 40% for the HIV-infected subcohort analysis, with \(X\) fairly small (e.g., \(X = 5000\) copies/mL) and \(T\) at least 18 months after diagnosis.
2. The lower 95% CI limit for VEVL\(_C\)\((T;X)\) > 40% for the full cohort analysis, with \(X\) fairly small (e.g., \(X = 5000\) copies/mL) and \(T\) the duration over which VE\(_C\) is assessed.
3. The efficacy results 1 and 2 are met for VECD\(_4\)\((T;X)\), with \(X = 350\) cells/mm\(^3\), with a less stringent threshold for the lower 95% lower 95% CI limit (e.g., at least 20%).
4. Inferences on VEVL\(_C\)\((T;X)\) and VECD\(_4\)\((T;X)\) are generally consistent across ranges for \(T\) and \(X\).

Note that analyses of either composite end point in the full cohort capture aggregate effects of vaccine to prevent infection and to prevent failure after infection. Consequently, if the estimate of VE\(_C\) is positive, then the efficacy thresholds listed above are more easily met for the full-cohort than subcohort analyses.

Furthermore, note that the exact efficacy thresholds (e.g., 40%) depend on specific characteristics of the populations used to guide the numerical study. For an actual trial, numerical exercises tailored to the study population and design would be used to obtain the exact thresholds.

**DISCUSSION**

If a prophylactic HIV vaccine confers limited protective efficacy against infection, we believe it should be considered for licensure. However, using an efficacy trial to identify such a vaccine is challenging, because of the need to monitor participants and their partners over the long term and because interpretation of long-term effects is made difficult by the use of ART. This article addresses the following question: What evidence from the trial will provide enough confidence about the vaccine’s benefits to justify licensure? Requiring demonstration of unequivocal clinical benefit to prevent disease and transmission before licensure is, in our opinion, overly cautious in an epidemic that is accruing 5 million new infections and 3 million deaths annually and are increasing [77], possibly delaying use of an effective vaccine. The delay could itself result in a large excess of HIV infections and deaths. The high human cost of delay, as well as the fairly high degree of understanding of biomarker end points and their relationship to HIV pathogenesis and transmission, argue that a vaccine showing positive and moderately durable effects on virus load, along with consistent positive effects on CD4-based and other biomarker end points merits licensure. Because of the uncertainties of basing decisions on surrogate end points without assurance of persistent clinically significant vaccine effects, early licensure must be accompanied by well-defined follow-up studies of HIV-infected participants and phase 4 epidemiological studies, to verify durable virological and immunological vaccine benefit, as well as ultimate clinical benefit [78]. A variety of phase 4 designs could contribute important information, such as partner studies that assess vaccine effects on rates and genotypes/phenotypes of transmitted HIV strains and community-based studies that assess the overall public health impact of a vaccination program [25]. A process for early vaccine licensure with subsequent confirmation of durable clinical efficacy may require the creation of new regulatory vehicles, which is being pursued by others [79] and is not addressed here.

In this article, we make a specific recommendation regarding the use of a surrogate end point as a coprimary end point along with HIV infection. We propose that virologic failure or treatment initiation can serve as the surrogate. Assessed over a sufficiently long period, this end point measures the durability of viremic control, which is a major determinant of disease and transmission rates. In addition to its value as a key surrogate for clinical events, the composite end point has an important ethical advantage in that the validity of its analysis is not compromised by treatment of HIV-infected participants. If HIV infection and the composite end point are used as coprimary end points, then the trial should be powered to detect a clinically significant magnitude of efficacy for either end point. In particular, the trial could be powered both for rejecting VE\(_C\) ≤ 30% and for rejecting VE\(_C\) ≤ 40%, because significance of either hypothesis test would imply a clinically significant vaccine effect. Note, however, that the appropriate null hypothesis for each efficacy parameter depends on the duration over which it is measured and, for VE\(_C\), also depends on the virologic failure threshold. Significant efficacy VE\(_C\) consistent across follow-up periods of 6–18 months after infection diagnosis and failure thresholds of 5000–55,000 copies/mL, may be required for concluding efficacy. In addition, significant efficacy may be required for both the intent-to-treat analysis of the randomized cohort and the subset analysis of the HIV-infected cohort, as well as these results for the composite end point based on CD4 cell count. Such consistency of evidence would robustly support that the vaccine is efficacious to ameliorate disease and transmission.

Use of numerical studies to calculate thresholds of vaccine effects on surrogate end points that reliably predict substantial vaccine effects on clinical end points, as recommended here, is limited by the data available for modeling and the uncertain long-term effects of vaccination on progression and transmissibility. Although there is risk in licensing a vaccine based mainly on surrogate end-point effects, we believe that the benefits outweigh the risks for populations not expected to increase risk behavior after vaccination; for such populations, the pub-
lic-health cost of licensing an ineffective vaccine is less than that associated with delaying licensure of an effective vaccine. Arguably, risk behavior would not increase for many populations, because the comprehensive prevention education that must accompany vaccine trials, their extended follow-up, and vaccine delivery programs [80] may itself lower risk behavior (e.g., large reductions in HIV incidence have attended vaccine trial activities in Thailand and Uganda). On the other hand, in some populations, risk behavior may increase after vaccination because of “disinhibition,” which could offset or reverse any benefits of vaccination programs [81]. Therefore, vaccines not showing clear efficacy to prevent HIV infection must be applied judiciously [80], and further studies are needed to evaluate licensability of vaccines in populations susceptible to postvaccination risk-behavior changes. In addition, it should be noted that licensure decisions based on protective efficacy against HIV acquisition also entail risk, because the efficacy could wane due to diminishing immunologic memory [82] or the evolving antigenicity of HIV, and mathematical models show that durability of efficacy is essential for a substantial population benefit [2, 83].

Modeling exercises suggest that a virus-suppressing vaccine may have its largest impact through population effects to reduce secondary HIV transmission [16, 75, 83]. Such a vaccine effect on transmissibility could vary by HIV strain, potentially yielding harm or benefit to a population. For example, on one hand, the use of a vaccine could skew prevalent HIV strains toward virulent, vaccine-resistant, and/or ART-resistant strains [84], or, on the other hand, toward mild and/or ART-sensitive strains. These considerations highlight the importance of carefully studying infecting strains and immune responses in efficacy trials to evaluate the possibility of selectively protective vaccine immunity [44, 70]. However, the information available from a phase 3 trial about vaccine effects on virus populations is limited, because only a few hundred trial participants seroconvert. Therefore, it is important that, after licensure of a vaccine, large phase 4 studies be conducted to monitor the changing patterns of HIV genotypes, phenotypes, and recombinants, in relation to immune responses that might predict resistance or susceptibility to the vaccine or ART. As multiple phase 3 and 4 HIV vaccine trials accrue, establishment of a cross-protocol registry for HIV-infected participants may help ensure adequate resistance monitoring. Because of the implementation of a vaccine or vaccine/treatment policy, collection of such data is needed for informing models that predict the population impact of the policy. The phase 4 and cross-protocol studies also would provide data on behavioral change that would importantly inform such simulation models. In addition, simulation models may be useful for directly interpreting vaccine effects on surrogate end points in terms of the predicted reduction in the basic reproduction number [13] or the predicted reduction in the number of new infections or AIDS deaths in a population during a time period.

Consideration of the composite end point illustrates that ART is not a barrier to the development of prophylactic HIV vaccines whose main benefits are on postinfection outcomes. Rather, development of vaccines and treatments is complementary, especially because seroconverters in efficacy trials can be entered into rollover protocols (possibly randomized) that evaluate the combined effect of prophylactic vaccine with various therapeutic interventions, potentially involving ARTs, therapeutic vaccines and cytokines [44, 85], and/or structured treatment interruptions [86]. Use of and adherence to standardized guidelines on initiation of ART and other therapeutic strategies is important for the success of both the vaccine trial and the rollover protocol.

APPENDIX

DETAILS OF THE NUMERICAL STUDY FOR IDENTIFYING THE MAGNITUDE OF ESTIMATED VACCINE EFFECTS ON THE SURROGATE END POINTS NEEDED TO REASONABLY INFER CLINICALLY SIGNIFICANT EFFICACY

The numerical study is based on the particular trial described in figure 2. The prediction equations described in Analysis of Surrogate End Points for Licensure Decisions of Candidate HIV Vaccines—Predicted VE(I(3)) = [1 - (1/2.39)^{AVL(3)}] × 100%; Predicted VE(I(3)) = [1 - [1 - (1 - Pr_{rel})^3]/Pr_{rel}] × 100%, where R = (1/2.39)^{AVL(3)}; and Predicted VE(I(3)) = VEVL(I(18);X) are modified to compute 95% prediction intervals for VE(I(3)) and VE(I(3)) that account for uncertainty sources (1)–(4) (see Analysis of Surrogate End Points for Licensure Decisions of Candidate HIV Vaccines). The calculations also incorporate the uncertainty in the relative-risk estimates 2.45 and 2.39, with published variance estimates 0.0205 and 0.0179, respectively [14, 43].

Uncertainty source (1) is accounted for by considering the variances of estimates of ΔVL(3) and VEVL(I(18);X) at the 3 specific analysis times described in figure 2. The variance for the ΔVL(3) end points depend on the number of HIV-infected subjects with virus load data and on the variance of the pre-ART virus load between 2 individuals, which we selected to be 0.36 on the basis of incident cohort data, assuming 2 measurements per subject [20]. The variance for the estimated VEVL(I(18);X) was determined by Monte Carlo simulations, assuming a constant hazard rate of the composite end point in each of the vaccine and placebo groups and a 50% failure rate by 18 months in the placebo group. Regarding uncertainty source (2), a simulation study suggests that selection bias is
unlikely to alter the mean initial pre-ART difference by more than 0.33 loga if VEa ≤40% [74]. The predictions were made robust to the selection bias by replacing the observed mean difference Estimated ΔVL(3) with the attenuated Estimated ΔVL(3) − 0.33. For VEVLc(18;X), robustness was built in by replacing the Estimated VEVLc(18;X) with 0.90 × Estimated VEVLc(18;X).

To address uncertainty source (3), a sensitivity analysis was conducted that assumes different degrees of reliability of the surrogate end-point effects as predictors of VEI(3) and VEP(84). Toward this goal, for predicting VEI(3) from ΔVL(3), we modified the original prediction equation to Predicted VEI(3) = [1 − (1/2.45) ΔVL(3) − 0.33] × (1 − f) × 100%, where f = b − (1 − p), p is the proportion of the vaccine effect on VEI(3) explained by the surrogate end-point effect to overpredict VEI(3) or VEP(84). The 95% prediction interval for VEp(84) given an observed Estimated VEVLc(18;X) is computed as follows: 1 − exp [−(1 − f) × log[1−0.90 × Estimated VEVLc(18;X)] + 1.96 × 0.90 × v2(1/2)], where b = 0.25, 0.50, 0.75, 1.0, reflecting a surrogate end point with “poor,” “moderate,” “good,” or “perfect” predictive reliability, and the value of b is chosen as 0.50 or 1.0, reflecting “substantial” or “extreme” bias of the surrogate end point effect to overpredict VEI(3) or VEP(84). For b = 0.50, 0.25, 0.125, 0.0 for p = 0.25, 0.50, 0.75, and 1.0, respectively, and, for b = 1.0, f = 0.75, 0.50, 0.25, 0.0 for p = 0.25, 0.50, 0.75, and 1.0, respectively. Uncertainty source (4) is accounted for through the number of HIV-infected participants with a minimum of 18 months follow-up after infection detection, which affects the precision of the surrogate parameter estimates that, in turn, affects the precision of the predictions.

Under the above set-up, the 95% prediction interval for VEI(3) given an observed Estimated ΔVL(3) is computed as follows: 1 − exp [−(1 − f) × log(2.45) × \{Estimated ΔVL(3) − 0.33\} + 1.96 × v2(1/2)], where b = 0.25, 0.50, 0.75, and 1.0, reflecting a surrogate end point with “poor,” “moderate,” “good,” or “perfect” predictive reliability, and the value of b is chosen as 0.50 or 1.0, reflecting “substantial” or “extreme” bias of the surrogate end point effect to overpredict VEI(3) or VEP(84). For b = 0.50, 0.25, 0.125, 0.0 for p = 0.25, 0.50, 0.75, and 1.0, respectively, and, for b = 1.0, f = 0.75, 0.50, 0.25, 0.0 for p = 0.25, 0.50, 0.75, and 1.0, respectively. Uncertainty source (4) is accounted for through the number of HIV-infected participants with a minimum of 18 months follow-up after infection detection, which affects the precision of the surrogate parameter estimates that, in turn, affects the precision of the predictions.

\[\text{Predicted VEI(3)} = (1 - \frac{\text{Estimated } \Delta VL(3) - 0.33}{2.45}) \times (1 - f) \times 100\%\]

\[\text{95% prediction interval for VEp(84)} = 1 - \exp\left[-(1-f) \times \log\left[1-0.90 \times \text{Estimated VEVLc(18;X)}\right] + 1.96 \times 0.90 \times \nu^{1/2}\right]\]

\[\text{95% prediction interval for VEI(3)} = 1 - \exp\left[-(1-f) \times \log\left(\frac{\text{Estimated } \Delta VL(3) - 0.33}{2.45}\right) + 1.96 \times \nu^{1/2}\right]\]

\[\nu = \text{variance estimate of } \log(2.45) \times \{\text{Estimated } \Delta VL(3) - 0.33\} \times 0.90\]

References


