MAJOR ARTICLE

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

Peter B. Gilbert,[1] Victor G. DeGruttola,[3] Michael G. Hudgens,[1] Steven G. Self,[1] Scott M. Hammer,[4] and Lawrence Corey[2]

[1]Statistical Center for HIV/AIDS Research and Prevention, Fred Hutchinson Cancer Research Center, and [2]Department of Microbiology, University of Washington, Seattle; [3]Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts; [4]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-

Received 11 September 2002; accepted 10 February 2003; electronically published 1 July 2003.

Presented in part: 14th International AIDS Conference, 7–12 July 2002, Barcelona, Spain (abstract MoPeD3643).

Financial support: National Institutes of Health (grants AI46703, AI24643, and AI054165-01 to P.B.G., M.G.H., and S.G.S.; grant AI51164 to V.G.D.; grants AI48013, AI46386, and AI42848 to S.M.H.; and grant AI45206 to L.C.).

Reprints or correspondence: Dr. Peter Gilbert, Statistical Center for HIV/AIDS Research and Prevention, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave. North, Seattle, WA, 98109 (pgilbert@scharp.org).

The Journal of Infectious Diseases 2003; 188:179–93

0022-1899/2003/18802-0001$15.00

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 $\geqslant$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, VE_S, 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 VE_P and VE_I that measure vaccine efficacy to reduce progression and infectiousness: VE_P is the percentage of reduction in the risk of AIDS or death by a time point well beyond HIV infection and VE_I 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 VE_P 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 VE_I 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 VE_S, this standard could be met by showing substantial VE_P and VE_I combined with demonstration that susceptibility to infection is not increased by vaccination. Because of the difficulties in demonstrating effects on VE_P and VE_I 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 VE_P and VE_I; such studies are important for understanding the surrogate vaccine effects. Furthermore, if vaccine effects on surrogate end points reliably predict VE_P and VE_I, 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

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

Vaccine efficacy parameters	Analyzed cohort	Time period for measuring end points
Clinical		
$VE_S(T)$: vaccine efficacy susceptibility		
Percentage of reduction in the risk of acquiring HIV infection by T months for vaccine vs. placebo recipients	Randomized cohort	Between entry and T months after entry (e.g., $T = 36$ months)
$VE_P(T)$: vaccine efficacy disease progression		
Percentage of reduction in the cumulative risk of progressing to AIDS or death by T months for vaccine vs. placebo recipients	Infected subcohort	Between infection diagnosis and T months after infection diagnosis (e.g., $T = 84$ months)
$VE_I(T)$: vaccine efficacy infectiousness		
Percentage of reduction in the per-contact probability of transmitting HIV to others at T months[a] for vaccine vs. placebo recipients	Infected subcohort	At T months after infection diagnosis (e.g., $T = 3$ months)
Surrogate		
$\Delta VL(T)$: vaccine efficacy pretreatment virus load		
Mean reduction in virus load measured before treatment at T months[b] for vaccine vs. placebo recipients	Infected subcohort	Between infection diagnosis and T months after infection diagnosis (e.g., $\Delta VL(3)$ = vaccine efficacy on initial pretreatment virus load)
$\Delta CD4(T)$ vaccine efficacy pretreatment CD4 cell count		
Mean reduction in CD4 cell count measured before treatment at T months[b] for vaccine vs. placebo recipients	Infected subcohort	Between infection diagnosis and T months after infection diagnosis (e.g., $T = 18$ months)
$VEVL_C(T;X)$: vaccine efficacy virologic failure or treatment initiation susceptibility		
Percentage of reduction in the cumulative risk of virologic failure (virus load $>X$ (e.g., $X > 55{,}000$ copies/mL) or treatment initiation (whichever occurs first) by T months for vaccine vs. placebo recipients	Infected subcohort	Between infection diagnosis and T months after infection diagnosis (e.g., $T = 18$ months)
	Randomized cohort	Between entry and T months after entry (e.g., $T = 36$ months)
$VECD4_C(T;X)$: vaccine efficacy CD4 cell count failure or treatment initiation susceptibility		
Percentage of reduction in the cumulative risk of CD4 failure (CD4 cell count $<X$; e.g., $X < 350$ copies/mm^3) or treatment initiation (whichever occurs first) by T months for vaccine vs. placebo recipients	Infected subcohort	Between infection diagnosis and T months after infection diagnosis (e.g., $T = 18$ months)
	Randomized cohort	Between entry and T months after entry (e.g., $T = 36$ months)

[a] Alternatively to defining $VE_I(T)$ based on infectious contacts occurring at T months, $VE_I(T)$ can be defined based on infectious contacts occurring during a time period T_1 to T months (e.g., 1–6 months after infection diagnosis).

[b] Alternatively to defining $\Delta VL(T)$ or $\Delta CD4(T)$ 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_1 to T months

AIDS and death [27–32] (analogous to VE_P). 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 observational studies showing 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,

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, $\Delta VL(T)$, 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 levels 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 T 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 $VEVL_C(T;X)$, 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 $VEVL_C(T;X)$, with $T = 18$ months, while emphasizing that all seroconverters should be monitored for at least 5 years and that $VEVL_C(T;X)$ be analyzed with T as large as feasible at each analysis time [e.g., update analyses of $VEVL_C(36,X)$ and $VEVL_C(60,X)$].

Figure 1 illustrates the analysis of $VEVL_C(18;X)$ 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, $VEVL_C(18;X)$ 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 $VEVL_C(18;X)$ 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 $VEVL_C(18;X)$ 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-

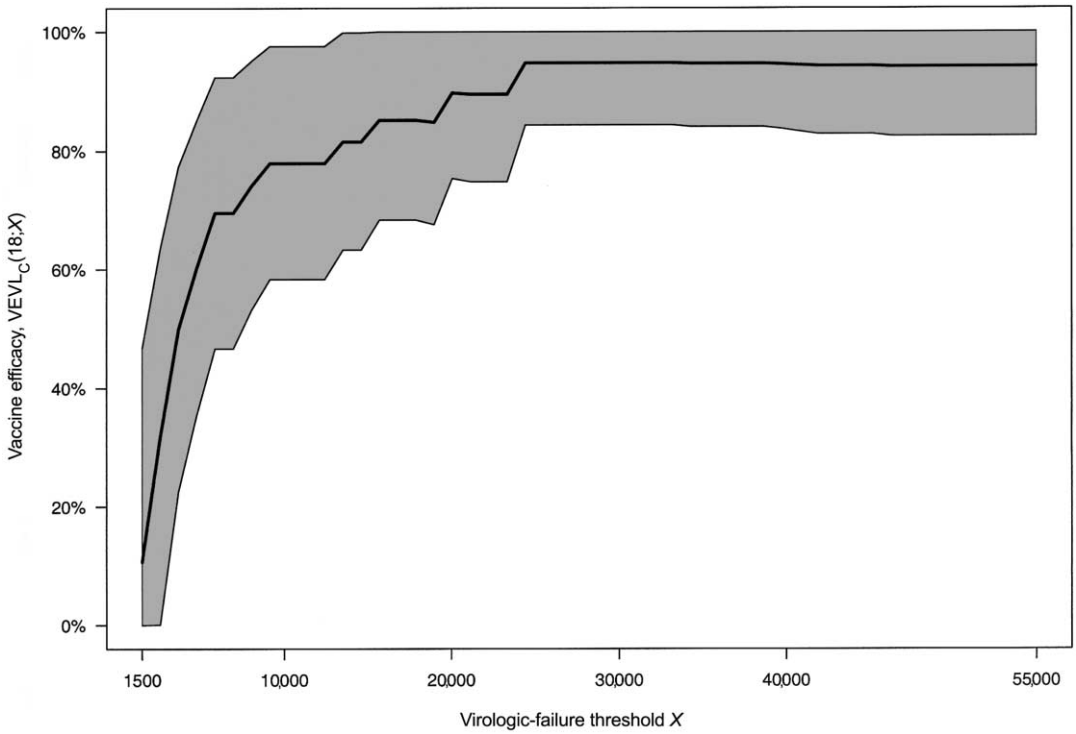

Figure 1. Analysis of $VEVL_C(18;X)$ (vaccine efficacy virologic failure >X or treatment initiation within 18 months) for a range of virologic failure thresholds X. For a vaccine trial data set simulated under design parameters given in figure 2, the bold line indicates the estimated $VEVL_C(18;X)$, plotted vs. the virologic-failure threshold X used in the end-point definition. For a given threshold X, $VEVL_C(18;X)$ is estimated by 1 − 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 $VEVL_C(18;X)$.

teria; 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 $VEVL_C(18;X)$ 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 $VEVL_C(18;X)$ applies to a population with treatment initiation guidelines and patterns similar to those observed in the trial. To complement $VEVL_C(18;X)$, we also considered the surrogate vaccine-efficacy parameter $\Delta VL(T)$ 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 $\Delta VL(T)$ applies to a population who would not use ART during the first T months after diagnosis of infection. $\Delta VL(T)$ 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 $\Delta VL(T)$. 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 $\Delta VL(T)$ 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-

ings in medications and treatment delivery infrastructure costs. Because of the complementary interpretations of the various surrogate end points, we suggest that all of them should be assessed, to provide a well-rounded picture of postinfection vaccine effects.

ANALYSIS OF SURROGATE END POINTS FOR LICENSURE DECISIONS OF CANDIDATE HIV VACCINES

The first 2 phase 3 trials of HIV vaccines have a single primary end point for efficacy: HIV infection [17, 18]. Consequently, support for licensure would derive mainly from analyses of VE_S. However, it is plausible that a vaccine strategy could only modestly reduce acquisition of infection but markedly ameliorate viremia after acquisition. Such a vaccine probably would provide considerable benefit in many countries; however, it would not qualify for licensure under current trial parameters that place HIV infection as the sole primary end point. Such strict interpretation of data could delay or prevent use of a product with significant therapeutic and transmission utility. On the other hand, the disadvantage of basing a licensure decision on an analysis of postinfection surrogate end points is the considerable uncertainty associated with the inference, stemming from the following 4 uncertainties.

1. Uncertainty in estimating the vaccine effects on the surrogate end points.
2. Potential statistical bias resulting from the fact that the analyzed groups are a select subset of the originally randomized groups [74, 75].
3. The fact that the vaccine effects on the surrogate end points are not fully validated as reliable predictors of vaccine effects on the clinical outcomes [34–39, 43, 68].
4. The dearth of data on the durability of the vaccine effects on the surrogate end points.

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_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 $VE_S > 30\%$, if the true VE_S is $\geqslant 60\%$ [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_S and the surrogate parameters for VE_P and VE_I. Each matrix divides assessment of vaccine effects into effects on the infection end point (VE_S) and effects on the postinfection end points (surrogates for VE_P/VE_I). 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_S, the "+," "−," and "·" decisions may be determined by whether the lower 95% CI limit for VE_S is >30%, the upper 95% CI limit is <0%, or otherwise, respectively [19]. Here, we developed decision criteria for the postinfection component VE_P/VE_I 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_P and VE_I. 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_S 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 $VE_I(3)$ and $VE_P(84)$ can be predicted from an estimate of $\Delta VL(3)$, as well as how well $VE_P(84)$ can be predicted from an estimate of $VEVL_C(18;X)$. The prediction of $VE_I(3)$ 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-$\log_{10}$ lower plasma virus load of the exposing partner. Specifically, $VE_I(3)$ is predicted as follows: Predicted $VE_I(3) = [1 - (1/2.45)^{\Delta VL(3)}] \times 100\%$. The prediction of $VE_P(84)$ 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-$\log_{10}$ lower initial plasma virus load (with adjustment for baseline CD4 cell count) [43]. This result leads to the following equation: Predicted $VE_P(84) = [[1 - \{[1 - (1 - PR_{84})^R]/PR_{84}\}]] \times 100\%$, where $R = (1/2.39)^{\Delta VL(3)}$ and PR_{84} is the progression rate in the placebo group by 84 months. The prediction of $VE_P(84)$ from $VEVL_C(18;X)$ is based on the following equation: Predicted

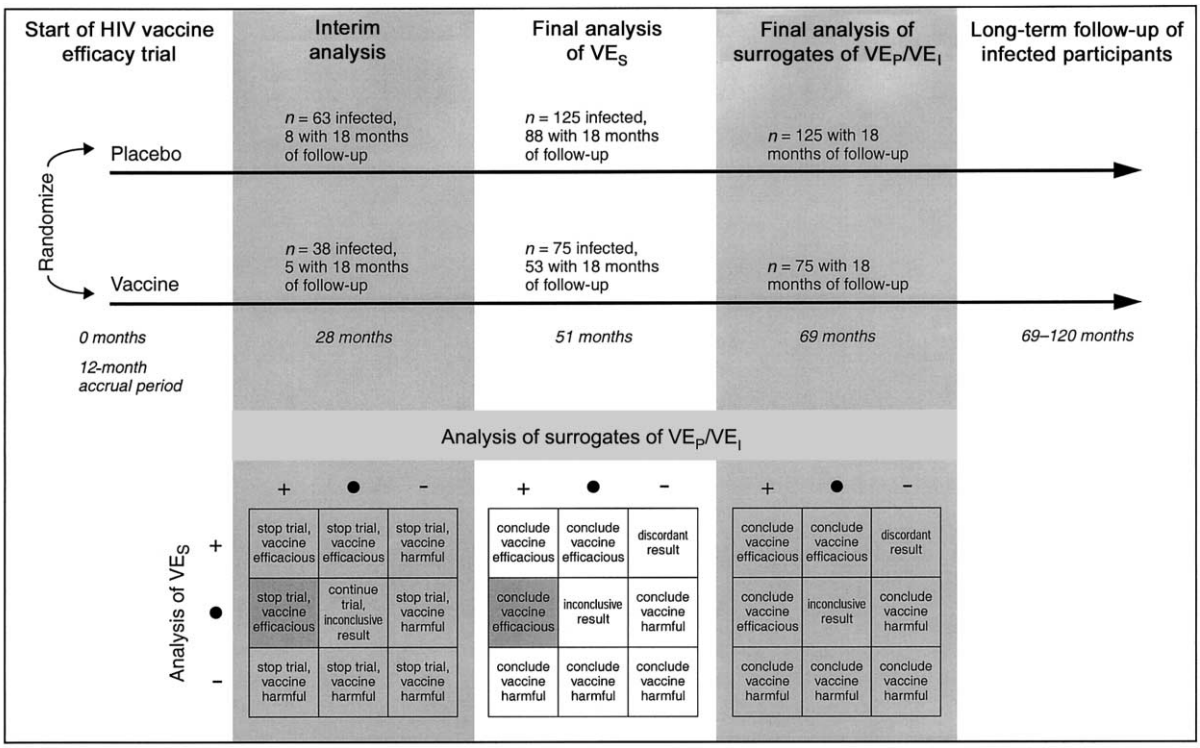

Figure 2. Schema and decision matrices for a prophylactic human immunodeficiency virus (HIV) vaccine efficacy trial. The top half of the figure depicts an example of an efficacy trial design. Accrual is uniform over 12 months, the interim analysis is triggered by the 63rd placebo infection and is expected at 28 months, the final analysis for VE_S (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 VE_P/VE_I (vaccine efficacy disease progression/vaccine efficacy infectiousness) is 18 months after the final analysis for VE_S. An estimated VE_S 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 VE_P/VE_I, given inconclusive evidence on VE_S? 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.

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

Data from the efficacy trial provide estimates of $\Delta VL(3)$, $VEVL_C(18;X)$, and PR_{84}, which can be substituted into the above equations to estimate Predicted $VE_I(3)$ and Predicted $VE_P(84)$. 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 VE_I and VE_P, 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 VE_I or VE_P ("poor," "moderate," "good," or "perfect") and (B) the extent to which unreliability of the surrogate end point leads to overprediction of VE_I or VE_P ("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 $VE_I(3)$ and $VE_P(84)$ 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 $VE_I(3)$ as a function of the lower 95% CI limit for $\Delta VL(3)$. Each panel represents a different combination of (A) and (B) defined above. Note that as the predictive reliability of the surrogate end point increases, a smaller effect on virus load is needed to infer a substantial $VE_I(3)$. 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 $VE_P(84)$ 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 $\Delta VL(3)$ (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 $\Delta VL(3)$ $\geqslant$1.3 $\log_{10}$ reliably predicts $VE_I(3)$ $\geqslant$40%, based on the lower 95% prediction limit for $VE_I(3)$. From this panel in figure 4, a lower 95% CI limit for $VEVL_C(18;X)$ $\geqslant$42% reliably predicts $VE_P(84)$ $\geqslant$40%. In addition, a lower 95% CI limit for $\Delta VL(3)$ $\geqslant$ 1.7 $\log_{10}$ reliably predicts $VE_P(84)$ $\geqslant$ 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.

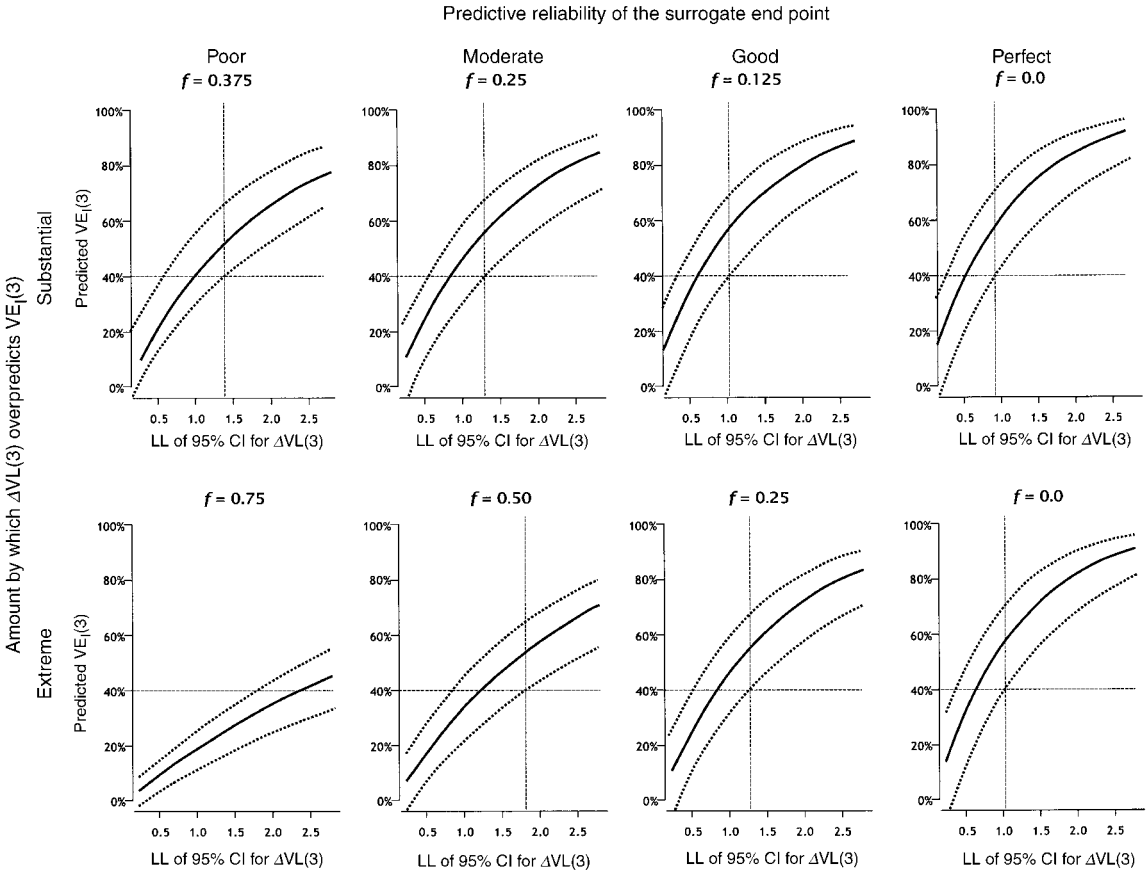

Figure 3. Predicted $VE_I(3)$ from the estimated $\Delta VL(3)$ (vaccine efficacy initial pretreatment virus load) at the final analysis of VE_S. Solid lines indicate $VE_I(3)$ (vaccine efficacy infectiousness) predicted from the lower limit (LL) of the 95% confidence interval (CI) for $\Delta VL(3)$, the vaccine efficacy to decrease initial pretreatment virus load; dotted lines are 95% prediction intervals for $VE_I(3)$. 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 $VE_I(3)$. If the surrogate end point is imperfect, then the surrogate vaccine effect may overpredict $VE_I(3)$. 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.

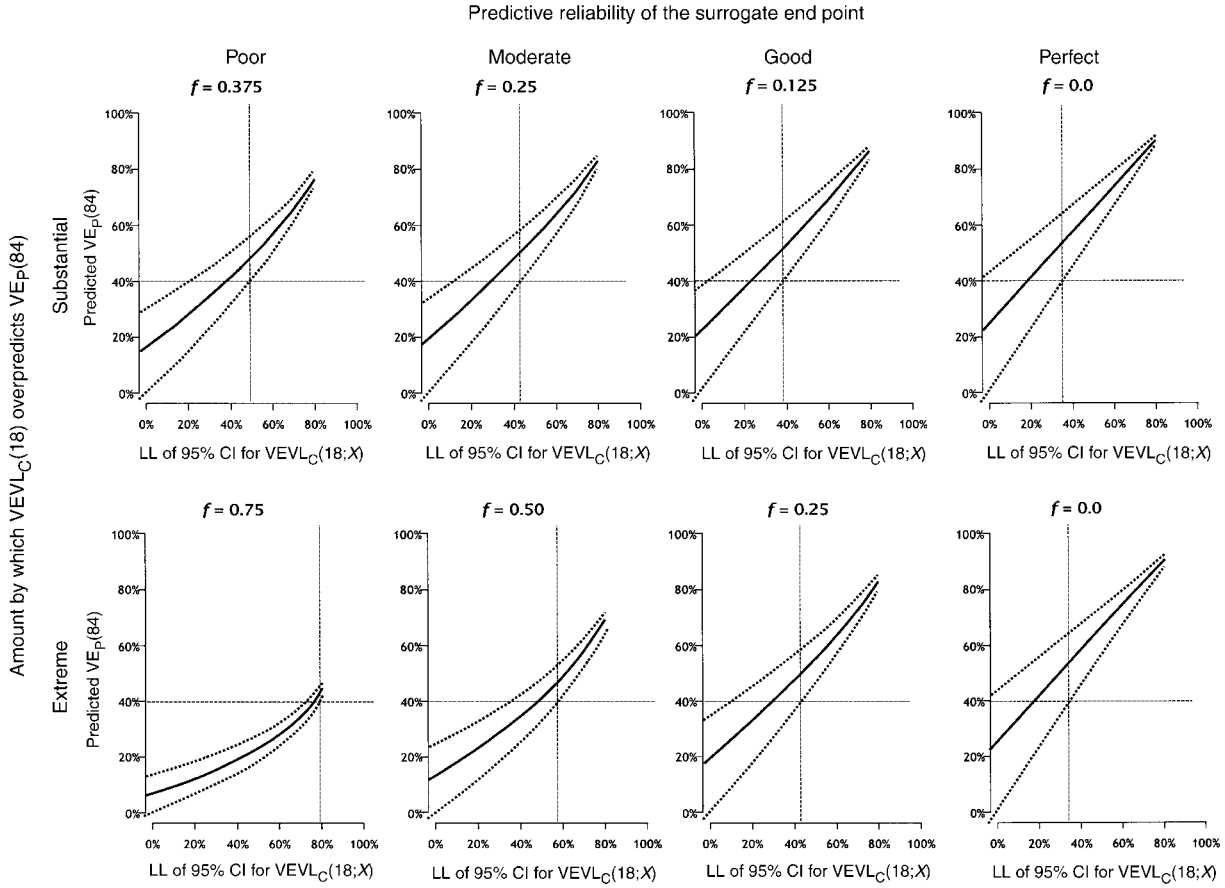

Figure 4. Predicted VE_P from the estimated $VEVL_C(18;X)$ (vaccine efficacy virologic failure >X or treatment initiation within 18 months) at the final analysis of VE_S. Solid lines indicate $VE_P(84)$ (vaccine efficacy disease progression within 84 months) predicted from the lower limit (LL) of the 95% confidence interval (CI) for $VEVL_C(18;X)$, 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 $VE_P(84)$. 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 $VE_P(84)$. If the surrogate end point is imperfect, then the surrogate vaccine effect may over-predict $VE_P(84)$. The predictions correct for an assumed bias to over-predict; 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.

On the basis of simulation experiments demonstrating the beneficial impact of using a vaccine with moderate VE_P and VE_I [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 VE_S, a vaccine showing inconclusive VE_S 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 $VEVL_C(18;X)$ >40%, plus consistent effects on CD4 cell counts. If this positive result occurs, the final analysis of the surrogates of VE_P/VE_I 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-

infected 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_S is assessed.
3. The efficacy results 1 and 2 are met for $VECD4_C(T;X)$, 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 $VECD4_C(T;X)$ 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_S 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, but largely prevents disease and secondary transmission, 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 $VE_S \leqslant 30\%$ and for rejecting $VE_C \leqslant 40\%$, 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 wielding 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 $VE_I(3) = [1 - (1/2.45)^{\Delta VL(3)}] \times 100\%$; Predicted $VE_P(84) = \{1 - [1 - (1 - PR_{84})^R]/PR_{84}\} \times 100\%$, where $R = (1/2.39)^{\Delta VL(3)}$; and Predicted $VE_P(84) = VEVL_C(18;X)$ are modified to compute 95% prediction intervals for $VE_I(3)$ and $VE_P(84)$ 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 0.0203 and 0.0179, respectively [14, 43].

Uncertainty source (1) is accounted for by considering the variances of estimates of $\Delta VL(3)$ and $VEVL_C(18;X)$ at the 3 specific analysis times described in figure 2. The variance for the $\Delta VL(3)$ 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 $VEVL_C(18;X)$ 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 $\log_{10}$ if $VE_S \leqslant 40\%$ [74]. The predictions were made robust to the selection bias by replacing the observed mean difference Estimated $\Delta VL(3)$ with the attenuated Estimated $\Delta VL(3) - 0.33$. For $VEVL_C(18;X)$, robustness was built in by replacing the Estimated $VEVL_C(18;X)$ with $0.90 \times$ Estimated $VEVL_C(18;X)$.

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 $VE_I(3)$ and $VE_P(84)$. Toward this goal, for predicting $VE_I(3)$ from $\Delta VL(3)$, we modified the original prediction equation to Predicted $VE_I(3) = [1 - (1/2.45)^{(\Delta VL(3) - 0.33) \times (1-f)}] \times 100\%$, where $f = b(1 - p)$, p is the proportion of the vaccine effect on $VE_I(3)$ explained by $\Delta VL(3)$ in relative-risk regression models [87, 88], and b is a bias factor between 0 and 1 that specifies the degree to which vaccine effects on $VE_I(3)$ through mechanisms other than the virus load end point attenuate the predicted $VE_I(3)$ toward 0%. A perfect surrogate end point has $p = 1$, yielding $f = 0$, and a surrogate that explains none of the vaccine effect has $p = 0$, yielding $f = b$. The smaller the p, the more likely the translation from $\Delta VL(3)$ to $VE_I(3) = [1 - (1/2.45)^{(\Delta VL(3) - 0.33)}] \times 100\%$ is incorrect, and $VE_I(3)$ could be higher or lower than predicted. To include robustness, we assumed that $VE_I(3)$ is, in fact, less than predicted, which is specified by setting $f > 0$. The constant b calibrates how much less, where $b = 0$ yields $f = 0$ and implies the predicted $VE_I(3)$ is the same as when the surrogate end point is perfect and $b = 1$ yields $f = 1 - p$ and represents the maximum plausible overprediction. Similar robustness calculations can be done to predict $VE_P(84)$ from ΔVL, yielding the modified equation Predicted $VE_P(84) = [[1 - \{[1 - (1 - PR_{84})^R] / PR_{84}\}]] \times 100\%$, with $R = (1/2.39)^{(\Delta VL(3) - 0.33) \times (1 - f)}$. In addition, the $VE_P(84)$ predicted from $VEVL_C(18;X)$ is $1 - [1 - 0.90 \times VEVL_C(18;X)]^{(1-f)}$. The value of p is chosen as 0.25, 0.50, 0.75, or 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 $VE_I(3)$ or $VE_P(84)$. For $b = 0.50$, f is 0.375, 0.25, 0.125, 0.0 for $p = 0.25$, 0.50, 0.75, and 1.0, respectively, and, for $b = 1.0$, f is 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 $VE_I(3)$ given an observed Estimated $\Delta VL(3)$ is computed as follows: $1 - \exp\{-(1 - f) \times \log(2.45) \times [\text{Estimated } \Delta VL(3) - 0.33] - 1.96 \times v^{1/2}]\}$, $1 - \exp\{-(1-f) \times \log(2.45) \times [\text{Estimated } \Delta VL(3) - 0.33] + 1.96 \times v^{1/2}]\}$, where v is the variance estimate of $\log(2.45) \times [\text{Estimated } \Delta VL(3) - 0.33]$, computed as $v = 0.0203 \times 0.36 + 0.0203 \times [\text{Estimated } \Delta VL(3) - 0.33]^2 + 0.36 \times [\log(2.45)]^2$. The 95% prediction interval for $VE_P(84)$ given an observed Estimated $VEVL_C(18;X)$ is computed as follows: $1 - \exp\{-(1 - f) \times [\log[1 - 0.90 \times \text{Estimated } VEVL_C(18;X)] + 1.96 \times 0.90 \times v^{1/2}]\}$, $1 - \exp\{-(1 - f) \times \log[1 - 0.90 \times \text{Estimated } VEVL_C(18;X)] - 1.96 \times 0.90 \times v^{1/2}]$, where v is the variance estimate of $\log[1 - 0.90 \times \text{Estimated } VEVL_C(18;X)]$, computed using Monte Carlo simulations. These formulas were used for creating figures 3 and 4. A similar formula was used for predicting $VE_P(84)$ from Estimated $\Delta VL(3)$. For these predictions, a constant hazard rate of AIDS/death was assumed for each group.

References

1. Anderson R, May R. Epidemiological parameters of HIV transmission. Nature **1988**; 333:514–9.
2. Anderson RM, Garnett GP. Low-efficacy HIV vaccines: potential for community-based intervention programs. Lancet **1996**; 348:1010–3.
3. Esparza J, Bhamarapravti N. Accelerating the development and future availability of HIV-1 vaccines: why, when, where, and how? Lancet **2000**; 355:2061–6.
4. Barouch DH, Santra S, Schmitz JE, et al. Control of viremia and prevention of clinical AIDS in rhesus monkeys by cytokine-augmented DNA vaccination. Science **2000**; 290:486–92.
5. Amara RR, Villinger F, Altman JD, et al. Control of a mucosal challenge and prevention of AIDS by a multiprotein DNA/MVA vaccine. Science **2001**; 292:69–74.
6. Rose NF, Marx PA, Luckay A, et al. An effective AIDS vaccine based on live attenuated vesicular stomatitis virus recombinants. Cell **2001**; 106:539–49.
7. Shiver JW, Fu T-M, Chen L, et al. Replication-incompetent adenoviral vaccine vector elicits effective anti-immunodeficiency virus immunity. Nature **2002**; 415:331–5.
8. Clements-Mann ML. Lessons for AIDS vaccine development from non-AIDS vaccines. AIDS Res Hum Retroviruses **1998**;14 (Suppl 3):S197–203.
9. Graham BS. Clinical trials of HIV vaccines. In: Kuiken C, Foley B, Hahn B, et al., eds. HIV Sequence Compendium 2000: theoretical biology and biophysics, group T-10. Los Alamos, NM: Los Alamos National Laboratory, **2000**:82–105.
10. Nabel GJ. Challenges and opportunities for development of an AIDS vaccine. Nature **2001**; 410:1002–7.
11. Shen X, Siliciano RF. Preventing AIDS but not HIV-1 infection with a DNA vaccine. Science **2000**; 290:463–5.
12. International AIDS Vaccine Initiative. State of current AIDS vaccine research. Available at: http://www.iavi.org/
13. Anderson RM, May RM. Infectious diseases of humans: dynamics and control. Oxford: Oxford University Press, **1991**.
14. Quinn TC, Wawer MJ, Sewankambo N, et al. Viral load and heterosexual transmission of human immunodeficiency virus type 1. Rakai Project Study Group. N Engl J Med **2000**; 342:921–9.
15. Gray RH, Wawer MJ, Brookmeyer R, et al. Probability of HIV-1 transmission per coital act in monogamous, heterosexual, HIV-1 discordant couples in Rakai, Uganda. Lancet **2001**; 357:1149–53.
16. Stover J, Walker N, Garnett GP, et al. Can we reverse the HIV/AIDS pandemic with an expanded response? Lancet **2002**; 360:73–7.
17. Choopanya K, Tappero J, Pitisuttithum P, et al. Ongoing follow-up of infecting drug users (IDUs) in the AIDSVAX B/E vaccine efficacy trial in Bangkok, Thailand [abstract WeOrD1300]. In: Program and abstracts of the 14th International AIDS Conference (Barcelona). Vol. 2. Stockholm: International AIDS Society, **2002**:149.
18. Francis DP, Gregory T, McElrath MJ, et al. Advancing AIDSVAX to

phase 3: safety, immunogenicity, and plans for phase 3. AIDS Res Hum Retroviruses **1998**; 14(Suppl 3):S325–31.
19. Rida W, Fast P, Hoff R, Fleming T. Intermediate-sized trials for the evaluation of HIV vaccine candidates: a workshop summary. J Acquir Immune Defic Syndr Hum Retrovirol **1997**; 16:195–203.
20. Schacker T, Collier AC, Hughes J, Shea T, Corey L. Biological and virologic characteristics of primary HIV infection. Ann Intern Med **1998**; 128: 613–20.
21. Longini IM, Hudgens MG, Halloran ME. Estimation of vaccine efficacy for both susceptibility to infection and reduction in infectiousness for prophylactic HIV vaccines with partner augmentation. In: Kaplan E, Brookmeyer R, eds. The quantitative evaluation of HIV prevention programs. New Haven, CT: Yale University Press, **2002**:241–59.
22. Mellors JW, Munoz A, Giorgi JV, et al. Plasma viral load and $CD4^+$ lymphocytes as prognostic markers of HIV-1 infection. Ann Intern Med **1997**; 126:946–54.
23. Lyles RH, Munoz A, Yamashita TE, et al. Natural history of human immunodeficiency virus type 1 viremia after seroconversion and proximal to AIDS in a large cohort of homosexual men. Multicenter AIDS Cohort Study. J Infect Dis **2000**; 181:872–80.
24. Longini IM, Datta S, Halloran ME. Measuring vaccine efficacy for both susceptibility to infection and reduction in infectiousness for prophylactic HIV-1 vaccines. J Acquir Immun Defic Syndr Hum Retrovirol **1996**; 13: 440–7.
25. Halloran ME, Struchiner CJ, Longini IM. Study designs for evaluating different efficacy and effectiveness aspects of vaccination. Am J Epidemiol **1997**; 146:789–803.
26. Falk LA, Ball LK. Current status and future trends in vaccine regulation: USA. Vaccine **2001**; 19:1567–72.
27. Mocroft A, Vella S, Benfield TL, et al. Changing patterns of mortality across Europe in patients infected with HIV-1. Lancet **1998**; 352:1725–30.
28. Palella FJ Jr, Delaney KM, Moorman AC, et al. Declining morbidity and mortality among patients with advanced human immunodeficiency virus infection. N Engl J Med **1998**; 338:853–60.
29. Vittinghoff E, Scheer S, O'Malley P, Colfax G, Holmberg SD, Buchbinder SP. Combination antiretroviral therapy and recent declines in AIDS incidence and mortality. J Infect Dis **1999**; 179:717–20.
30. Detels R, Munoz A, McFarlane G, et al. Effectiveness of potent antiretroviral therapy on time to AIDS and death in men with known HIV infection duration. Multicenter AIDS Cohort Study Investigators. JAMA **1998**; 280:1497–503.
31. Hoyert DL, Kochanek KD, Murphy SL. Deaths: final data for 1997. Nat Vital Stat Rep **1999**; 47:1–104.
32. Centers for Disease Control and Prevention. Update: AIDS—United States, 2000. MMWR Morb Mortal Wkly Rep **2002**; 51:592–5.
33. Hammer SM, Squires KE, Hughes MD, et al. A controlled trial of two nucleoside analogues plus indinavir in persons with human immunodeficiency virus infection and CD4 cell counts of 200 per cubic millimeter or less. N Engl J Med **1997**; 337:725–33.
34. Fleming TR, DeMets DL. Surrogate end points in clinical trials: are we being misled? Ann Intern Med **1996**; 125:605–13.
35. Fleming TR. Evaluating therapeutic interventions (with Discussion and Rejoinder). Stat Science **1992**; 7:428–56.
36. DeGruttola V, Wulfsohn M, Fischl M, Tsiatis A. Modeling the relationship between survival and $CD4^+$ lymphocytes in patients with AIDS and AIDS-related complex. J Acquire Immun Defic Syndr **1993**; 6: 359–65.
37. Fleming TR. Surrogate markers in AIDS and cancer trials. Stat Med **1994**; 13:1423–35.
38. HIV Surrogate Marker Collaborative Group. Human immunodeficiency virus type 1 RNA level and CD4 count as prognostic markers and surrogate endpoints: a meta-analysis. AIDS Res Hum Retroviruses **2000**; 16:1123–33.
39. DeGruttola VG, Clax P, DeMets DL, et al. Considerations in the evaluation of surrogate endpoints in clinical trials. Cont Clin Trials **2001**;22: 485–502.
40. Goodman GE, Schaffer S, Bankson DD, Hughes MP, Omenn GS. Predictors of serum selenium in cigarette smokers and the lack of association with lung and prostate cancer risk. Cancer Epidemiol Biomarkers Prev **2001**; 10:1069–76.
41. O'Brien TR, Blattner WA, Waters D, et al. Serum HIV-1 RNA levels and time to development of AIDS in the Multicenter Hemophilia Cohort Study. JAMA **1996**; 276:105–10.
42. Mellors JW, Rinaldo CR Jr, Gupta P, White RM, Todd JA, Kingsley LA. Prognosis in HIV-1 infection predicted by the quantity of virus in plasma. Science **1996**; 272:1167–70.
43. Albert JM, Ioannidis JPA, Reichelderfer P, et al. Statistical issues for HIV surrogate endpoints: point/counterpoint. Stat Med **1998**; 17:2435–62.
44. Report from a meeting of the WHO-UNAIDS Vaccine Advisory Committee Geneva, 21–23 February 2000. Approaches to the development of broadly protective HIV vaccines: challenges posed by genetic, biological and antigenic variability of HIV-1. AIDS **2001**; 15:W1–25.
45. Lukashov VV, Goudsmit J, Paxton WA. The genetic diversity of HIV-1 and its implications for vaccine development. In: Wong-Staal F, Gallo RC, eds. AIDS vaccine research. New York: Marcel Dekker, **2002**:93–120.
46. Borrow P, Lewicki H, Hahn BH, Shaw GM, Oldstone MB. Antiviral pressure exerted by HIV-1–specific cytotoxic T lymphocytes (CTLs) during primary infection demonstrated by rapid selection of CTL escape virus. Nat Med **1997**; 3:205–11.
47. Price DA, Goulder PJ, Klenerman P, et al. Positive selection of HIV-1 cytotoxic T lymphocyte escape variants during primary infection. Proc Natl Acad Sci USA **1997**; 94:1890–5.
48. Goulder PJ, Phillips RE, Colbert RA, et al. Late escape from an immunodominant cytotoxic T-lymphocyte response associated with progression to AIDS. Nat Med **1997**; 3:212–7.
49. Goulder PJ, Brander C, Tang Y, et al. Evolution and transmission of stable CTL escape mutations in HIV infection. Nature **2001**; 412:334–8.
50. Barouch DH, Kunstman J, Kuroda MJ, et al. Eventual AIDS vaccine failure in a rhesus monkey by viral escape from cytotoxic T lymphocytes. Nature **2002**; 415:335–9.
51. Murray JS, Elashoff MR, Iacono-Connors LC, Cvetkovich TA, Struble KA. The use of plasma HIV RNA as a study endpoint in efficacy trials of antiretroviral drugs. AIDS **1999**; 13:797–804.
52. Staszewski S, DeMasi R, Hill AM, Dawson D. HIV-1 RNA, CD4 cell count and the risk of progression to AIDS and death during treatment with HIV-1 reverse transcriptase inhibitors. AIDS **1998**; 12:1991–7.
53. Katzenstein DA, Hammer SM, Hughes MD, et al. The relationship of virologic and immunologic markers to clinical outcomes after nucleoside therapy in HIV-infected adults with 200 to 500 CD4 cells per cubic millimeter. N Engl J Med **1996**; 335:1091–8.
54. Jordan R, Gold L, Cummins C, Hyde C. Systematic review and meta-analysis of evidence for increasing numbers of drugs in antiretroviral combination therapy. BMJ **2002**; 324:1–10.
55. Egger M, May M, Genevieve C, et al. Prognosis of HIV-1–infected patients starting highly active antiretroviral therapy: a collaborative analysis of prospective studies. Lancet **2002**; 360:119–28.
56. Gilbert PB, DeGruttola V, Hammer SM, Kuritzkes DR. Virologic and regimen termination surrogate end points in AIDS clinical trials. JAMA **2001**; 285:777–84.
57. Gulick R, Mellors J, Havlir D, et al. Treatment with indinavir, zidovudine, and lamivudine in adults with human immunodeficiency virus infection and prior antiretroviral therapy. N Engl J Med **1997**; 337: 734–9.
58. Collier AC, Coombs RW, Schoenfeld DA, et al. Treatment of human immunodeficiency virus infection with saquinavir, zidovudine, and zalcitabine. AIDS Clinical Trials Group. N Engl J Med **1996**; 334:1011–7.
59. Panel on Clinical Practices for Treatment of HIV Infection, Department of Health and Human Services. Guidelines for the use of antiretroviral agents in HIV-infected adults and adolescents. 4 February **2002**. Available at: http://www.aidsinfo.nih.gov/guidelines/
60. Durant J, Clevenbergh P, Halfon P, et al. Drug-resistance genotyping in HIV-1 therapy: the VIRADAPT randomized controlled trial. Lancet **1999**; 353:2195–9.
61. Hirsch MS, Brun-Vezinet F, D'Aquila RT, et al. Antiretroviral drug

resistance testing in adult HIV-1 infection: recommendations of an International AIDS Society–USA panel. JAMA **2000**; 283:2417–26.
62. World Health Organization. Scaling up antiretroviral therapy in resource limited settings: guidelines for a public health approach, executive summary. April **2002**. Available at: http://www.who.int/HIV_AIDS/first.html.
63. Geise R, Maenza J, Celum CL. Clinical challenges and diagnostic approaches to recognizing acute human immunodeficiency virus infection. Am J Med **2001**; 111:237–8.
64. Guenter D, Esparza J, Macklin R. Ethical considerations in international HIV vaccine trials: summary of a consultative process conducted by the Joint United Nations Programme on HIV/AIDS (UNAIDS). J Med Ethics **2000**; 26:37–43.
65. Rotnitzky A, Robins JM, Scharfstein DO. Semiparametric regression for repeated outcomes with nonignorable nonresponse. J Amer Stat Assoc **1998**; 93:1321–39.
66. Liang KY, Zeger SL. Longitudinal data analysis using the generalized linear model. Biometrika **1986**; 73:13–22.
67. Verbeke G, Molenberghs G. Linear mixed models for longitudinal data. New York: Springer, **2000**.
68. Hughes MD, Daniels MJ, Fischl MA, Kim S, Schooley RT. CD4 cell count as a surrogate endpoint in HIV clinical trials: a meta-analysis of studies of the AIDS Clinical Trials Group. AIDS **1998**; 12:1823–32.
69. Pitt J, Brambilla D, Reichelderfer P, et al. Maternal immunologic and virologic risk factors for infant human immunodeficiency virus type 1 infection: findings from the Women and Infants Transmission Study. J Infect Dis **1997**; 175:567–75.
70. Gilbert P, Self S, Rao M, Naficy A, Clemens J. Sieve analysis: methods for assessing from vaccine trial data how vaccine efficacy depends on genotypic and phenotypic pathogen variation. J Clin Epidemiol **2001**; 54:68–85.
71. Sterling TR, Vlahov D, Astemborski J, Hoover DR, Margolick JB, Quinn TC. Initial plasma HIV-1 RNA levels and progression to AIDS in women and men. N Engl J Med **2001**; 344:720–5.
72. Hogan CM, Hammer SM. Host determinants in HIV infection and disease. II. Genetic factors and implications for antiretroviral therapeutics. Ann Intern Med **2001**; 134:978–96.
73. Tang J, Shelton B, Makhatadze NJ, et al. Distribution of chemokine receptor CCR2 and CCR5 genotypes and their relative contribution to human immunodeficiency virus type 1 (HIV-1) seroconversion, early HIV-1 RNA concentration in plasma, and later disease progression. J Virol **2002**; 76:662–72.
74. Hudgens MG, Hoering A, Self SG. On the analysis of viral load endpoints in HIV vaccine trials. Stat Med (in press).
75. Gilbert PB, Bosch RJ, Hudgens MG. Sensitivity analysis for the assessment of causal vaccine effects on viral load in HIV vaccine trials. Biometrics (in press).
76. Barth-Jones DC, Longini IM. Determining optimal vaccination policy for HIV vaccines: a dynamic simulation model for the evaluation of vaccine policy. In: Andersen JG, Katzper M, eds. Proceedings of the International Conference on Health Sciences Simulation 2002. San Antonio: 2002 Western Multiconference, **2002**:63–80.
77. UNAIDS. Report on the global HIV/AIDS epidemic 2002. Available at: http://www.unaids.org/barcelona/presskit/report.html.
78. Orenstein WA, Bernier RH, Hinman AR. Assessing vaccine efficacy in the field. Epidemiol Rev **1988**; 10:212–41.
79. Isbell MT, Widdus R. Actions to avoid potential regulatory delays for vaccines against HIV/AIDS. Available at: http://www.iavi.org.
80. Future access to HIV vaccines: report from a WHO-UNAIDS consultation, Geneva, 2–3 October 2000. AIDS **2001**; 15:W27–44.
81. Velasco-Hernandez JX, Gershengorn HB, Blower SM. Could widespread usage of combination antiretroviral therapy eradicate HIV epidemics? Lancet Infect Dis **2002**; 2:487–93.
82. Kaul R, Rowland-Jones SL, Kimani J, et al. Late seroconversion in HIV-resistant Nairobi prostitutes despite pre-existing HIV-specific $CD8^+$ responses. J Clin Invest **2001**; 107:341–9.
83. Anderson RM, Swinton J, Garnett GP. Potential impact of low efficacy HIV-1 vaccines in populations with high rates of infection. Proc R Soc Lond B **1995**; 261:147–51.
84. O'Connor D, Allen T, Watkins DI. Vaccination with CTL epitopes that escape: an alternative approach to HIV vaccine development? Immunol Lett **2001**; 79:77–84.
85. Hel Z, Venzon D, Poudyal M, et al. Viremia control following antiretroviral treatment and therapeutic immunization during primary SIV_{251} infection of macaques. Nat Med **2000**; 6:1140–6.
86. Altfeld M, Rosenberg E, Shankarappa R, et al. Cellular immune responses and viral diversity in individuals treated during acute and early HIV-1 infection. J Exp Med **2001**; 193:169–180.
87. Freedman LS, Graubard BI, Schatzkin A. Statistical validation of intermediate endpoints for chronic diseases. Stat Med **1992**; 11:167–78.
88. Lin DY, Fleming TR, DeGruttola V. Estimating the proportion of treatment effect explained by a surrogate marker. Stat Med **1997**; 16:1515–27.