Some Design Issues in Trials of Microbicides for the Prevention of HIV Infection

Thomas R. Fleming and Barbra A. Richardson
Department of Biostatistics, University of Washington, Seattle

(See the editorial commentary by Padian, on pages 663–5.)

Trials for the prevention of human immunodeficiency virus (HIV) infection that evaluate microbicides provide significant design challenges. Three of these design issues deserve more careful consideration. The first issue relates to the benefits of using both blinded and unblinded control groups when the placebo regimen may not be inert and when the effectiveness of an intervention heavily depends on behavioral, as well as biological, factors. The second issue relates to the strength of evidence required for regulatory approval for the marketing of drugs and biologics when only a single pivotal phase 3 clinical trial has provided such evidence. The third issue relates to the appropriate next step after the completion of phase 1 trials, as well as the specific merits of conducting phase 2b screening trials that assess the effects on the same clinical efficacy end point that will be the primary end point in a phase 3 trial. The issues considered in microbicide trials for the prevention of HIV infection are also of importance in many other clinical scenarios.

In 2002, >11,000 new cases of HIV infection occurred among adults each day, with the majority of these cases transmitted through heterosexual contact [1]. Although sexual transmission of HIV can be avoided through consistent use of male condoms [2], it is difficult for many women to negotiate condom use by their male partners [3–5]. Topical vaginal microbicides are products designed to be an alternative method for the prevention of sexual transmission of HIV and other disease pathogens [3–6].

Clinical trials of candidate vaginal microbicides provide significant design and implementation challenges. Among these challenges is the need to achieve high levels of adherence to experimental microbicide regimens and to available standard-of-care interventions, such as condom use, as well as the need to achieve high levels of retention during at least 1–2 years of follow-up after randomization. The importance of such issues has been discussed frequently [7, 8]. Hence, in this manuscript, we will consider 3 important design issues that have not been so widely discussed. These issues relate to (1) the benefits of implementing both blinded and unblinded control groups, (2) the strength of evidence required to receive regulatory approval for the marketing of drugs and biologics when evidence for safety and efficacy essentially is provided by a single trial, and (3) the choice of the appropriate next step after completion of phase 1 trials, as well as the specific merits of conducting phase 2b screening trials. Trials of vaginal microbicides will be used to motivate and illustrate these issues. Informed consent was obtained from all participants in the studies described or from their parents or guardians, and human experimentation guidelines of the US Department of Health and Human Services were followed in the conduct of clinical research.

BLINDING AND THE CHOICE OF CONTROL GROUPS

For trials that have subjective end points, such as the level of pain, nausea, or breathlessness, or end points that are based on the time to a major clinical inter-
vvention, such as hospitalization, bias can occur if a study participant’s treatment group is known to the study participant, the care giver, or the evaluator of the end points of the trial. In an effort to obtain unbiased estimates of benefit and risk, a randomized placebo-controlled clinical trial is often viewed as the method that is the reference standard. In the following discussion, the “efficacy” of a regimen will be defined as the effect that would be estimated in a blinded comparison with an inert placebo. In such a trial, the level of adherence to the experimental regimen, which will influence efficacy, should represent a high level of what is achievable in a real-world setting.

Even with these benefits of blinding, many trials are conducted with unblinded control regimens. There are several factors that motivate such an approach [8]. For example, either the experimental and control regimens may not be of a similar nature or they may have obvious and very different side-effect profiles, making it impractical or unethical to achieve blinding.

Trials of vaginal microbicides provide additional motivation to consider the use of unblinded control regimens. First, when an end point is quite objective (e.g., HIV infection), there is less risk for assessment bias than in trials that have end points that are subjective (e.g., pain relief).

Second, estimates of the efficacy of a microbicide that are obtained by comparing the microbicide with a placebo control gel are unbiased only if the placebo gel is inert against the disease pathogen of interest. However, vaginal microbicideresearch remains hindered by the inability to determine whether the placebo gels used in clinical trials truly are inert. Potential mechanisms of action that could cause the placebo gels used in vaginal microbicide trials to alter the risk of HIV acquisition include lubricating and/or physical barrier effects; the presence, in the placebo, of preservatives that may themselves have microbialicidal effects; alteration of the vaginal microflora by the placebo; and decreased concentrations of HIV in the female genital tract after ejaculation, as a result of the presence of a volume of placebo. A potentially protective “placebo” is the Replens gel (Columbia Laboratories) used in the Joint United Nations Programme on HIV/AIDS trial of nonoxynol-9 [9].

Third, inclusion of an unblinded “no-gel” group (i.e., a “condom-only” group) allows for assessment of the real-world “effectiveness” of a vaginal microbicide. The risk of HIV infection will be influenced by the level of protection provided when a vaginal microbicide is used; the level of adherence to use of the vaginal microbicide; the level of adherence to ancillary interventions, such as condoms; and the level of risk behaviors. Hence, “effectiveness,” which is defined as the reduction in the risk of HIV infection achieved by implementation of the experimental regimen in a real-world unblinded manner, is influenced by both biological and behavioral factors. Because study participants may vary their use of other protective interventions, such as condoms, and because they may also vary their sexual behaviors when they know that they are using a microbicide (versus when they know that they are not using a microbicide), comparison of subjects who are following the microbicide regimen with subjects in the unblinded no-gel group is needed to incorporate all elements required to achieve an assessment of real-world effectiveness.

Figure 1 presents the potential mechanisms of action of a microbicide intervention. As shown in figure 1A, comparison of the active microbicide regimen with a placebo gel would reveal the antimicrobial effects of the regimen and would also provide an unbiased estimate of efficacy if the placebo gel is inert. In turn, as shown in figure 1B, comparison of the microbicide regimen with an unblinded no-gel control addresses real-world effectiveness by taking into account (1) the antimicrobial effects of the microbicide, (2) other protective effects of the microbicide that might also be carried by the placebo, such as physical barrier effects and lubrication effects, and (3) effects resulting from the changes in risk behavior that may occur with the use of a microbicide product.

Figure 2 shows the insights that would be obtained from a trial that provides a comparison of the active microbicide regimen with both a placebo gel control arm and an unblinded no-gel control arm. For each of the 6 scenarios shown in figure 2, the annual incidence of HIV infection is presented for the active microbicide, placebo gel, and unblinded no-gel arms of the trial. In scenario 1a, the microbicide clearly would not be effective. In contrast, in scenario 2a, the microbicide would be very effective, with a 33% relative reduction in the incidence of HIV infection provided through its antimicrobial effect. Estimates of effectiveness and efficacy would be the same in this scenario, because the annual incidence of HIV infection among subjects in the placebo gel and unblinded no-gel control groups is the same. In scenarios 2b and 2c, the estimated efficacy

A Randomize __ Active Microbicide ~ Antimicrobial effects

B Randomize __ Active Microbicide Unblinded Control ~ Antimicrobial effects

~Physical Barrier Effects ~Effects on ~Lubrication Effects Risk Behavior ~Other Effects

Figure 1. Potential effects of a microbicide regimen that can be detected when a placebo group (A) or an unblinded control group (B) is used.
REQUIRED STRENGTH OF EVIDENCE

Regulatory authorities, such as the US Food and Drug Administration (FDA), have long recognized the wisdom of requiring confirmatory trials in the assessment of the efficacy and safety of drugs and biologics used for treatment and prevention. Approvals in a given clinical indication usually have been based on evidence provided by ≥2 adequate and well-controlled trials. However, when trials have evaluated effects on mortality or irreversible morbidity end points, such as stroke, loss of vision, or HIV infection, the FDA has considered evidence from a single trial to be adequate for approval if that evidence is “robust and compelling”. This is particularly true when uncommonly large or resource-intensive trials are required or when the disease under consideration is rare.

Given that the strength of evidence for declaring a single trial to be “positive” usually requires that a false-positive error rate of 0.025 be maintained (i.e., that a 2-sided P value of <.05 or, more accurately, a 1-sided P value of <.025 be obtained), the FDA often has viewed evidence of the effect of treatment on a major clinical end point to be “compelling,” when provided by only a single trial, if the 1-sided P value is bounded above by a value in the range of .0005–.005 (i.e., achieving a strength of evidence in the range of 1.5–2 positive trials; to be specific, .0006 and .004 are .025 raised to the 2.0 and 1.5 powers, respectively) [11, 12]. Of course, the final decision regarding approval would depend on several factors, including effects on safety measures and secondary end points, issues related to quality of the conduct of the trial, and relevant evidence from related external trials.

In a phase 3 trial designed to provide the sole evidence for efficacy of a microbicide regimen, meeting this increased strength of evidence would, in turn, require substantial increases in sample sizes and corresponding increases in study resources and duration. For example, consider a phase 3 trial designed to have high power to detect a 33% reduction in the risk of HIV infection, where time to HIV infection is analyzed using the Cox proportional hazards model. In the trial, for each pairwise comparison of an active regimen versus a control regimen, the number of events, L, that would be required to provide 90% power to achieve the strength of evidence of 2 positive trials, rather than 1 positive trial. (The sample size of the trial is, in turn, calculated as L divided by the fraction of study participants who have an “event”.)

Figure 3 shows the point estimates of efficacy that would be required to achieve various levels of strength of evidence, in trials for which L = 256 or L = 409. If L = 409, estimates of efficacy of 17.5% and 24.1% would be needed to achieve 1-sided P values of <.025 and <.0025, respectively.
Reduction in Risk of HIV Infection, %

**Scenario #1:** Target: 1-Sided $P < .025$; 256 End points

<table>
<thead>
<tr>
<th>0%</th>
<th>17.5%</th>
<th>21.5%</th>
<th>24%</th>
<th>27%</th>
<th>29.5%</th>
<th>33%</th>
</tr>
</thead>
<tbody>
<tr>
<td>.025</td>
<td>.0025</td>
<td>.0005</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Scenario #2:** Target: 1-Sided $P < .0025$; 409 End points

Figure 3. Efficacy estimates required to achieve various levels of strength of evidence.

With the insights from the scenarios presented in figure 2, it might be proposed that a single trial would serve as the basis for registration if comparisons of both the active microbicide with placebo and the active microbicide with the unblinded no-gel control regimen yielded a 1-sided $P$ value of $<.025$, with at least 1 of these comparisons achieving a 1-sided level of significance in the range of $P < .0025 - P < .0005$. Of course, as previously noted, the final decision regarding approval would depend on several factors, including effects on safety measures and secondary end points, as well as on the quality of the conduct of the trial.

In a trial providing the targeted $L = 409$ end points, there would be high probability of meeting these registration criteria in scenario 2a and, likely, also in scenarios 2b and 1b in figure 2. On the other hand, the microbicide likely should not be approved in scenarios 1a and 1c, because of lack of efficacy (although the placebo gel might be viewed as beneficial in scenario 1c), or in scenario 2c, where implementation of a regimen with 0% effectiveness would be problematic. It is important to emphasize that the availability of information from the unblinded control regimen could reduce the strength of evidence for benefit obtained from the placebo comparison, (e.g., in scenario 2c), but it could also increase such evidence (e.g., in scenario 1b).

**THE NEXT STEP AFTER A PHASE 1 TRIAL: A PHASE 2, PHASE 2B, OR PHASE 3 TRIAL?**

After completion of a phase 1 trial that has successfully established that the experimental intervention is worthy of additional study, determination of the proper next step in the clinical evaluation often involves considerable controversy. Traditionally, the next step would be to conduct ≥1 phase 2 trials. Such trials typically would include 50–200 participants, who often are drawn from a population similar or identical to the population that eventually would be enrolled in the phase 3 trial.

Phase 2 trials provide many valuable insights. Broader safety information is among the most important of these insights. Often, phase 2 trials also are designed to evaluate the dose, schedule, or mode of delivery of a drug. In addition, phase 2 trials can provide insights that will be useful in improving the quality of conduct of the phase 3 trial. For example, for vaginal microbicide regimens, these insights could relate to the ability of study sites to achieve timely enrollment of participants, approaches to provide high levels of adherence to the experimental microbicide regimen, insights regarding the delivery of the standard of care (such as the required level of intensity of counseling about condom use), and procedures that will ensure that the phase 3 trial will achieve high levels of retention.

Before embarking on an expensive phase 3 trial, one should have established considerable plausibility that the experimental intervention truly has a favorable benefit-to-risk profile. Usually, such plausibility also would be established in a phase 2 trial, in which there is confirmation of the effects on a biological marker through which intended clinical benefit is expected to be mediated. Unfortunately, in some scenarios, such as the use of vaginal microbicide regimens for the prevention of HIV infection, biological markers are not available. For such scenarios, what options exist to establish plausibility of efficacy? Moving directly to a phase 3 trial involving thousands of participants per arm, even if early phase 2-type safety assessments were to be conducted in the early stage of the trial, would be a very risky strategy that could result in very inefficient use of time and resources.

The phase 2b trial provides an intriguing option in this setting [14]. In such a trial, a screening evaluation of the experimental intervention is obtained by directly studying the effect of the intervention on the targeted clinical end point (e.g., for the scenario involving use of a microbicide, the effect on the incidence of HIV infection would be studied). The number of events in the phase 2b trial, $L^*$, that would be required for an efficient strategy could be specified by targeting approximately one-fourth to one-third of the number of events, $L$, that will be obtained in the phase 3 trial. It might be expected that efficiency would be achieved by setting $L^*$ to be approximately one-fourth to one-third of $L$, because a screening trial should be sufficiently large to provide reliable leads and, yet, should be much less costly than a phase 3 trial that will provide the definitive assessment of efficacy and safety. (Unpublished computer simulations have confirmed the efficiency [defined as minimizing the total number of events required to develop a compelling case for licensure] of this phase 2b design strategy. Hence, for the situation discussed in the Required Strength of Evidence section, in which the phase 3 microbicide trial would have $L = 409$ events for each comparison of active microbicide group versus control group, the phase 2b trial might have approximately $L^* = 100$ events for each pairwise comparison. The Appendix provides the details for the design of a phase 2b trial of microbicides, including an illustration of guidelines for determining whether to proceed to a phase 3 trial after completion of the phase 2b trial.
An alternative approach to an $L^*$-event phase 2b trial would be to conduct an $L$-event phase 2/3 trial that involves an early phase 2–type safety assessment, followed by an interim analysis of efficacy when $L^*$ events have occurred. Although this approach would allow all patients to be included in the “phase 3” component of the trial, the sponsor would not have access to the interim safety and efficacy data. Such data would only be available to a data monitoring committee [15], which would be guided by continuation criteria that were prespecified by the sponsor. Hence, such a trial truly is a phase 3 trial by design.

ILLUSTRATIONS OF TRIALS

Successful use of a phase 2b trial. The phase 2b screening trial is illustrated by the development of 5-fluorouracil (5-FU) plus levamisole, the only chemotherapy regimen to receive FDA approval for use in the treatment of stage-3 adjuvant colon cancer. In this setting, $\sim 50\%$ of control patients will experience recurrence of disease and death within 5 years after initiation of treatment.

The North Central Cancer Treatment Group conducted a phase 2b trial that evaluated the use of levamisole alone, as well as the combination of 5-FU plus levamisole [16]. This screening trial provided approximately $L^* = 90$ primary end points (i.e., deaths) per pairwise comparison with the unblinded no-treatment control arm. The results provided encouraging estimates of a 33% reduction in the mortality rate (figure 4). A confirmatory phase 3 trial was conducted by the National Institutes of Health Cancer Intergroup and provided approximately $L = 300$ deaths for each pairwise comparison with the control arm [17].

The Cancer Intergroup Trial (CIT) illustrated that it is possible to achieve timely enrollment of participants in a confirmatory trial, even when encouraging evidence was provided by a phase 2b trial conducted among subjects with a life-threatening disease, for whom lack of any approved treatments provided an urgent unmet need. The results of the CIT, as shown in figure 4, confirmed that the use of 5-FU plus levamisole reduced the death rate by 33% and led to the prompt regulatory approval of this regimen. Of interest, the 2 trials shown in figure 4 also illustrate that phase 2b trials can yield false-positive leads. Specifically, in the CIT, the regimen that involved the use of levamisole alone was found to provide no effect on survival. Hence, these trials illustrate that it is both possible and important to conduct confirmatory trials in a timely manner when phase 2b trials provide encouraging but not compelling evidence of efficacy.

**A phase 2b trial providing compelling evidence of benefit.** A phase 2b trial with $\sim 100$ primary end points can provide evidence of benefit that is sufficiently compelling to indicate that a subsequent phase 3 trial would not be necessary. An illustration of this is provided by the recent HIV Network for Prevention Trials (HIVNET) 012 trial, which established that a regimen involving single doses of nevirapine (given to the mother during labor and given to the infant after delivery) would provide an $\sim 50\%$ reduction in the rate of mother-to-child transmission of HIV, relative to a short-course regimen involving zidovudine (given to the mother during labor and delivery and given to the infant for 1 week after delivery) (see figure 5). At 14–16 weeks of age, 65 of 302 infants (Kaplan-Meier estimate, 25.1%) in the zidovudine arm and 37 of 307 infants (Kaplan-Meier estimate, 13.1%) in the nevirapine arm had HIV infection (1-sided $P = .0003$) [18]. In developing countries, this regimen involving single-dose nevirapine has now become widely implemented for the prevention of mother-to-child transmission of HIV.

**DISCUSSION**

In this manuscript, we have considered 3 important issues associated with the design of clinical trials. Although the relevance of these issues has been investigated and illustrated with regard to the use of vaginal microbicides for the prevention of heterosexual transmission of HIV, these issues are also of importance in many other clinical settings.

The first issue is associated with the merits of conducting randomized trials that include both blinded and unblinded...
control groups—in particular, in situations such as trials of vaginal microbicides, for which there remains substantial uncertainty about whether a placebo regimen is inert. Furthermore, when prevention of transmission of HIV depends not only on the level of protection provided when a prevention regimen is used but, also, on behavioral factors that influence risk behavior, it becomes very important to assess effectiveness as well as efficacy [19]. The comparison of a microbicidal regimen with an unblinded standard-of-care regimen provides this overall estimate of effectiveness. Contrary to the beliefs of some investigators [20], this estimate of effectiveness does not need to be adjusted according to the differences in the level of risk behaviors between subjects in the intervention and unblinded control arms, because such differences in behavior likely are an intrinsic part of the real-world effectiveness of the microbicidal regimen. Although efficacy and effectiveness might be underestimated in such clinical trials, because adherence to the microbicidal could be less than that in the real world after it has been established to have a favorable benefit-to-risk profile, the intensive oversight that participants experience in a clinical trial provides an enhanced level of adherence to potentially “correct” this hypothetical bias.

The use of unblinded control regimens would also provide a particularly appealing approach when there are ethical concerns that arise from potentially harmful effects of the placebo. Such harmful effects could be of a biological nature, such as induction of epithelial disruption in vaginal microbicidal trials. However, they could also be of a behavioral nature. Specifically, suppose that study participants are randomized to follow either an active microbicidal regimen or a blinded control regimen, to evaluate the outcome of HIV infection. If participants tend to reduce their adherence to effective standard-of-care regimens, such as use of condoms, because of perceived protection from the randomized intervention, the placebo participants could have a higher average risk of HIV infection than if they had not participated in the trial.

A second key issue is related to the required strength of evidence for regulatory approval when evidence is provided by only a single pivotal trial. The FDA has been flexible in allowing single pivotal trials when effects on major clinical end points, such as mortality, are evaluated—in particular, when trials are very resource intensive [11]. In such settings, the FDA typically has indicated that the single registrational trial must provide results that are “robust and compelling,” rather than specify a P value that corresponds to a targeted strength of evidence. This is very appropriate, given that regulatory approval decisions would depend on several factors, including effects on safety measures and secondary end points, quality of trial conduct, and relevant external results.

Sponsors would be prudent, when there is a single pivotal trial, to have high statistical power to achieve levels of significance in the range of .0025–.005, (1-sided P values), because irregularities can arise even in the best-planned and carefully conducted trials. An illustration is provided by the pivotal trial that evaluated Xigris (Eli Lilly), the first agent approved for use for individuals with severe sepsis. Even though this trial involved acute unmet need, and even though P values for the effect of Xigris on survival were lower than a 1-sided P value of .025 in this single pivotal trial, the FDA Anti-Infective Drugs Advisory Committee had a split 10–10 vote for approval because of several irregularities in that trial [12].

The final key design challenge considered in this manuscript is related to the options for the next step after completion of phase 1 clinical trials—in particular, when there was no biological marker available to assess plausibility of efficacy in a phase 2 clinical trial. Proceeding from a small phase 1 trial to a major phase 3 study would be a very risky strategy without first establishing effects on biological markers and, often, without having fully adequate insights about the optimal design of the phase 3 trial or the ability of the infrastructure of the clinical trials network to achieve high standards for quality-of-trial conduct. The phase 2b intermediate trial provides a measured next step to efficiently screen out ineffective regimens, to provide reliable leads regarding promising interventions, and to provide compelling evidence of benefit for highly effective therapies. The phase 2b trial also provides important insights about trial design and infrastructure preparedness if a phase 3 trial would be required. With the phase 2b trial as a supportive trial for registrational purposes, the phase 3 trial, if required, may only need to provide the traditional strength of evidence of a single positive trial.

In conclusion, when planning the next step after a phase 1 trial, the goal is not specifically to initiate the phase 3 trial(s) as soon as possible; rather, the goal should be to complete, as soon as possible, the trial(s) that provide “robust and compelling” evidence that the experimental intervention has a favorable benefit-to-risk profile.
APPENDIX

THE DESIGN OF A PHASE 2B SCREENING TRIAL: AN ILLUSTRATION

A phase 2b screening trial of microbicides for the prevention of HIV infection is illustrated by a major study sponsored by the Division of AIDS (DAIDS) of the National Institutes of Health and designed by the HIV Prevention Trials Network (HPTN). The trial, known as “HPTN 035,” is evaluating 2 candidate vaginal microbicides, BufferGel (ReProtect) and 0.5% PRO 2000/5 (P) gel (Indevus Pharmaceutical), and it includes a blinded placebo-gel control arm and an unblinded no-gel control arm. Through an extensive protocol development process, the DAIDS and the HPTN decided that the phase 2b trial would be the proper step with which to follow the completed phase 1 trials of BufferGel and 0.5% PRO 2000/5 (P) gel, each of which involved <100 participants. The DAIDS-assembled External Scientific Review Committee endorsed the judgment that use of the phase 2b screening trial design would allow an efficient strategy of research toward establishment of safe and effective vaginal microbicide regimens that can be marketed with regulatory approval.

GOALS OF THE PHASE 2B SCREENING TRIAL

The phase 2b trial will provide important insights about the safety of the 2 microbicide products and will provide screening evidence about efficacy and effectiveness. To be specific, it will provide evidence that suggests that an experimental microbicide regimen (1) is not adequately effective and should be discarded in its current formulation in this indication; or (2) is plausibly efficacious and should be evaluated more definitively in a subsequent phase 3 clinical trial; or (3) is efficacious, with strength of evidence of a single positive trial (for which 1-sided sequential phase 3 clinical trial; or (3) is efficacious, with strength of evidence about efficacy and effectiveness. To be specific, it will provide evidence that suggests that an experimental microbicide regimen that would be both clinically relevant as well as plausible to achieve. Hence, by use of estimates of efficacy obtained from data analyses that use the Cox proportional hazards model, the goal would be to distinguish regimens that provide no benefit from regimens that provide a reduction of ≥33% in HIV seroincidence.

On the basis of discussions with the Food and Drug Administration, a false-positive error rate of 0.0025, which, on a log scale, is approximately midway between the false-positive error rates associated with the strength of evidence from 1 and 2 traditional phase 3 studies (0.025 and 0.000625, respectively), would be used for hypothesis testing if a single phase 3 microbicide trial were being considered for registration. Such a trial would require L = 409 events for each pairwise comparison of intervention regimen versus control regimen. With this targeted number of events, the log relative risk would be estimated with an SE of 0.099. In turn, one would need an estimated reduction of ~17.5% in the HIV infection incidence rate to achieve a strength of evidence with 1-sided P = .025 and an estimated 24.1% reduction to achieve the definitive target of a strength of evidence with a 1-sided P value of .0025 (figure 3 and figure A1).

In contrast, as illustrated in figure A1, a phase 2b screening trial with one-fourth this amount of information (i.e., L' = 100), will have twice the SE in the estimates of the log relative risk. The primary analysis of the phase 2b trial will be formally based on a 4-category decision guideline.

Note that these categories should not be interpreted as providing strict decision rules but, rather, as guidelines that will be factored into a broader scientific assessment of the benefit to the risk profile of the microbicide regimen. This broader assessment will include consideration of safety, secondary efficacy end points, and relevant information external to this trial. Specifically, the decision guidelines for this trial are as follows:

1. If the estimated effect of a candidate microbicide is ≤15.3%, exclude the candidate microbicide from additional testing in its current formulation.
2. If the estimated effect is >15.3% but is ≤33%, consider the product to be plausibly effective and to merit further evaluation in a separate phase 3 trial with a false-positive error rate in the range of 0.0025–0.025, depending on evidence regarding other factors, such as safety, secondary end points, and measures of quality of trial conduct. In planning the design of the phase 3 trial, this phase 2b screening trial (and other information that emerges during the conduct of this trial) will provide important insights regarding:
   a. The estimated benefit to the risk profile of the
microbicide regimen;
b. Possible refinements in the dose and/or schedule of the microbicide regimen;
c. Approaches to improve adherence to the microbicide regimen;
d. Approaches to improve adherence to standard-of-care interventions, such as condom use;
e. Approaches to achieving timely accrual of study participants at centers with established ability to provide high-quality implementation of the trial and high-quality data; and
f. Reassessment of the choice of primary and secondary end points that will be clinically relevant and sensitive to the intended effects of the experimental intervention, as well as reassessment of the necessary sample size to provide acceptable statistical power.

3. If the estimated effect is between 33% and 43.6%, consider the product to be effective with a strength of evidence of at least a single phase 3 study (false-positive error rate, 0.025). A confirmatory trial, with a false-positive error rate of 0.025, likely would be required.

4. If the estimated effect is >43.6%, consider the product to be effective with the strength of evidence of at least one-and-a-half phase 3 studies (false-positive error rate, 0.0025).

The operating characteristics for this 4-category decision guideline are as follows:

1. The false-positive error rate of the screening procedure is low. If a candidate microbicide truly is ineffective (0% effect), there is only an 18% chance of proceeding to a phase 3 trial designed with a false-positive error rate in the range of 0.0025–0.025, a 2.2% chance of proceeding to a phase 3 trial designed with a false-positive error rate of 0.025, and only a 0.25% chance of declaring the microbicide to be effective.

2. The false-negative error rate of the screening procedure is low. If a candidate product truly reduces risk by 33%, there is only a 12.5% chance of discarding the product from additional testing, a 37.5% chance of proceeding to a phase 3 trial designed to have a false-positive error rate in the range of 0.0025–0.025, a 30.1% chance of proceeding to a phase 3 trial designed to have a false-positive error rate of 0.025, and a 19.9% chance of declaring the microbicide to be effective, (i.e., that 1-sided $P<0.0025$ would be achieved).

3. The power to provide a definitive result for a very effective candidate product is high. If a product truly reduces risk by 50%, then there is a 72.2% chance that the product would be declared effective (i.e., that 1-sided $P<0.0025$ would be achieved) and only a 0.49% chance that it would be discarded from additional testing.

Table A1 summarizes the probabilities of each possible decision for a variety of true effect levels. Probabilities were calculated using the approximate asymptotic normality of the estimate of the natural logarithm of the relative risk [13, chapters 4 and 8].

In the HPTN 035 phase 2b trial, decision criteria for future steps, including potential regulatory approval, would factor in the strength of evidence relative to both the placebo control arm and the unblinded no-gel control arm, as noted in the

Table A1. Probabilities of study outcomes, according to various levels of the true effect.

<table>
<thead>
<tr>
<th>TE, %</th>
<th>Of ≤15.3%</th>
<th>Of &gt;15.3% but ≤33%</th>
<th>Of &gt;33% but ≤43.6%</th>
<th>Of &gt;43.6%</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.792</td>
<td>0.183</td>
<td>0.022</td>
<td>0.0025</td>
</tr>
<tr>
<td>5</td>
<td>0.713</td>
<td>0.243</td>
<td>0.038</td>
<td>0.005</td>
</tr>
<tr>
<td>10</td>
<td>0.617</td>
<td>0.309</td>
<td>0.063</td>
<td>0.011</td>
</tr>
<tr>
<td>15</td>
<td>0.507</td>
<td>0.371</td>
<td>0.100</td>
<td>0.022</td>
</tr>
<tr>
<td>15.3</td>
<td>0.500</td>
<td>0.375</td>
<td>0.102</td>
<td>0.023</td>
</tr>
<tr>
<td>20</td>
<td>0.390</td>
<td>0.418</td>
<td>0.149</td>
<td>0.043</td>
</tr>
<tr>
<td>25</td>
<td>0.276</td>
<td>0.434</td>
<td>0.209</td>
<td>0.081</td>
</tr>
<tr>
<td>30</td>
<td>0.175</td>
<td>0.410</td>
<td>0.270</td>
<td>0.145</td>
</tr>
<tr>
<td>33</td>
<td>0.125</td>
<td>0.375</td>
<td>0.301</td>
<td>0.199</td>
</tr>
<tr>
<td>35</td>
<td>0.097</td>
<td>0.344</td>
<td>0.316</td>
<td>0.243</td>
</tr>
<tr>
<td>40</td>
<td>0.046</td>
<td>0.249</td>
<td>0.325</td>
<td>0.381</td>
</tr>
<tr>
<td>43.6</td>
<td>0.023</td>
<td>0.176</td>
<td>0.301</td>
<td>0.500</td>
</tr>
<tr>
<td>45</td>
<td>0.017</td>
<td>0.150</td>
<td>0.284</td>
<td>0.549</td>
</tr>
<tr>
<td>50</td>
<td>0.005</td>
<td>0.071</td>
<td>0.202</td>
<td>0.722</td>
</tr>
<tr>
<td>55</td>
<td>0.001</td>
<td>0.025</td>
<td>0.109</td>
<td>0.866</td>
</tr>
<tr>
<td>60</td>
<td>&lt;0.001</td>
<td>0.006</td>
<td>0.040</td>
<td>0.954</td>
</tr>
</tbody>
</table>

NOTE. TE, true effect (i.e., true percentage reduction in the rate of HIV infection provided by the experimental regimen).
discussion of the phase 3 trial in the Required Strength of Evidence section.

References