

# The Design of Prophylactic Trials for HIV: The Case of Microbicides

Zena A. Stein,<sup>1,2,3</sup> Landon Myer,<sup>1</sup> and Mervyn Susser<sup>1,2</sup>

Vaginal microbicides are topical applications used to help women protect themselves from sexually transmitted infection. Consideration of their role as a public health measure against human immunodeficiency virus (HIV) in particular began in the late 1980s.<sup>1</sup> By then it had become clear to many in the HIV field that advising women to insist that their partners use condoms was seldom realistic; it assumes a power that few of the women most in need possess. The underlying aim of the vaginal application was to provide a woman with a means of protecting herself without requiring her male partner to participate or even to know.

By February 2002, plans for at least five trials of candidate microbicides were approaching final form. At the time of this writing, none of these has entered phase 3, the final test in the field trial itself.<sup>2</sup> It seems to us that the design strategies of those studies closest to implementation are in some respects deficient. Thus, the moment is timely for raising epidemiologic concerns. In what follows we address two issues: one is the choice of controls, and the other has to do with problems and benefits of contemporaneous trials at multiple sites.

Randomized controlled trials (RCTs) of a new pharmaceutical product are mounted only after preliminary work (including tests on laboratory animals) has supported their likely efficacy. Phase 1 is conducted in human subjects primarily to ensure safety and to guide dosage. Safety is further assessed in phase 2, and the appropriate trial procedures are devised and developed in the "at-risk" population itself. In phase 3, investigators can begin to evaluate the product with continued

monitoring for safety and effectiveness in the field. In the United States, each step is monitored by the Food and Drug Administration (FDA).

## Control Groups

Although several aspects of design for microbicide trials have received close attention, the critical choice of a control arm still engenders contention. Two distinct types of controls have been proposed. One is the standard of choice, a double-blind placebo arm in which neither the participants nor the researchers are permitted to know the treatment assignment. The placebo should be an inert substance indistinguishable from the experimental product in physical characteristics.

In several of the new trials proposed, however, an additional control arm receives only condoms together with counseling on their use. In an editorial accompanying a recently published microbicide trial,<sup>3</sup> at least one experienced statistician approved a design in which a single control arm consisted of condoms without placebo.<sup>4</sup> We have encountered two grounds for the use of a "condom-only" arm: first, to ensure a determinate result; and second, to enable extrapolation to the population at large (the "real world"). In what follows, we voice our doubts about whether this tactic can be relied on to fulfill either of these intended purposes.

Concern about the selection of appropriate controls first arose in 2000 in the wake of the UNAIDS COL-1492 trial results.<sup>5</sup> In an untoward result, HIV incidence appeared slightly higher among subjects assigned to the experimental gel (and condoms) than among those assigned to the placebo (and condoms). The interpretation of this result remains uncertain. To explain it, some researchers (including authors of a World Health Organization [WHO] report<sup>6</sup>) have suggested that the experimental gel may not necessarily have caused harm. Instead, the placebo itself, not being the "inert" product it was assumed to be, could have had a preventive effect on HIV. To circumvent this design hazard, some measure to gauge whether the placebo was indeed inert would perhaps help.

---

**Editors' note:** An invited commentary on this article appears on page 83.

From the <sup>1</sup>Department of Epidemiology, Mailman School of Public Health, and <sup>2</sup>Gertrude H. Sergievsky Center, Columbia University; and the <sup>3</sup>HIV Center for Clinical and Behavioral Studies, New York State Psychiatric Institute, New York, NY.

Address correspondence to: Zena Stein, Sergievsky Center, Columbia University, 630 West 168 Street, Box 16, New York, NY 10032; sengupt@sergievsky.cpmc.columbia.edu

Submitted 23 September 2002; final version accepted 17 October 2002.

Copyright © 2002 by Lippincott Williams & Wilkins, Inc.

It is apparently to this end that the several impending American-sponsored trials (although so far not the British trial) have adopted the additional "condom-only" control arm. Thus, the proposed trial design incorporates dual controls: a placebo arm and a "condom-only" arm. In effectiveness trials, these are to be compared with one or more test products, the putative microbicides.

In evaluating the effectiveness of an experimental product, a design that adds new arms has a substantial impact on the time, resources, fieldwork and quantity of data in studies that are already very large. The implications of such an approach in testing microbicide effect therefore require careful scrutiny, not least with respect to inference.

Table 1 presents a set of the potential outcomes of testing a hypothetical experimental microbicide in a randomized trial of the kind proposed. Relative risks of HIV incidence in the microbicide arm are compared with those in the placebo and the "condom-only" arms. In rows A, B and C, the relation of the putative microbicides to both placebo and "condom-only" arms is the same. Interpretation is straightforward. The placebo arm provides the more credible reference for microbicide effect. At best, the "condom-only" arm supplies secondary confirmation.

In all the remaining rows of the Table (D to I), the relative effect of the putative microbicide differs depending on the control comparison. Each of these rows allows a discomfiting number of interpretations, some more and some less plausible, but no one of them secure. No great difficulty resides in interpreting the standard comparisons of putative microbicide *vs* placebo. But the third "condom-only" arm is a metaphoric joker. Addition of an unblinded group compromises the neutrality between arms with respect to behavior conferred by randomization.

Of particular note are rows G and H, in which the putative microbicide appears either equivalent to or worse than placebo (although both are better than "condom only"). In the COL-1492 trial (with no "condom-only" comparison) the result was indeterminate: either the placebo was more effective than the microbicide, or the microbicide was harmful. It is this disconcerting result that fueled calls for the inclusion of a "condom-only" arm in future microbicide trials. We turn now to consider whether such a step would serve a determinative result.

Because of the open nature of trial assignment in comparisons of a topical agent *vs* "condom only," outcomes could very well reflect differences in behavior rather than interventions. Across an open trial, reports of behavior and especially reports of sensitive sexual activity could vary with trial assignment. Thus, retrospective reports of sexual behavior and of condom use could be of questionable validity. Variability in actual behavior, in reported behavior and in interaction be-

tween these could yield three uncontrolled sources of variation between the participant groups across the arms of an open trial. We see no way in which the potentially confounding effects of such sources of variability can be adequately controlled. In comparisons across properly blinded groups, such doubts should not arise.

A second argument for the use of a "condom-only" arm assumes that the trial comparison involving such an arm will better represent the parent population untouched by intervention. Supposedly, this would measure the protective effect of microbicides on women's risk of HIV infection in the "real world." Randomized controlled trials do not aspire to represent experiences and effectiveness in the "real world." In the ideal, such trials test the effectiveness of a particular intervention by comparisons under arbitrary conditions created to conform to the imposed requirements of the research and differing only according to the intervention under test. Departures from *et ceteris paribus* (literally, all other things being equal) introduce the biases and confounding that randomization and blinding, the essential features of randomized control trials, are intended to prevent.

In a trial, two things act selectively and virtually preclude true representation of a specified population at large. One is the ethical requirement to enroll consenting volunteers, and the other is variability in readiness to comply with differing forms of intervention. Moreover, the generalizability of any clinical trial result will be problematic for reasons aside from the nature of the control arm. Under the conditions of research and of the "real world," measures of effect are bound to be different: first, behavioral changes are to be expected after the introduction of such new types of intervention as the microbicide; second, a gulf separates the rigor and demands of properly implemented clinical trials on the one hand and the delivery of health services on the other.

Strictly, the goal of phase 3 microbicide trials is to test whether a microbicidal product lessens the chances of HIV transmission. We conclude that the inclusion of a "condom-only" arm will not contribute to this goal, and could detract from it. Should any experimental microbicide appear ineffective in comparison with the best available placebo, then the experimental product is unlikely to be effective for widespread HIV prevention. Given limited resources and the tremendous urgency of identifying an effective microbicide for HIV prevention, the additional cost and time associated with a "condom-only" control arm is difficult to justify.

## Multiple Trials

The second issue concerns the multiple contemporaneous trials of microbicides that we anticipate will be entering the field in 2003.<sup>2</sup> This is a welcome development; the

**TABLE 1. Hypothetical Outcomes and Relative Risks (RR) of a Randomized Controlled Trial Comparing an Experimental Microbicide with Placebo and to "Condom Only"\*†**

	Relative Risk (RR) of HIV infection		Interpretation†
	vs Placebo	vs Condom Only	
A	RR < 1.0	RR < 1.0	Microbicide is effective
B	RR > 1.0	RR > 1.0	Microbicide is harmful
C	RR = 1.0	RR = 1.0	Microbicide has no impact on risk
D	RR < 1.0	RR = 1.0	Microbicide, preventive compared with placebo, is no better than "condom only"
E	RR < 1.0	RR > 1.0	Microbicide prevents HIV infection compared with placebo, but is less effective than "condom only"
F	RR > 1.0	RR = 1.0	Microbicide, worse than placebo, is equivalent to "condom only"
G	RR > 1.0	RR < 1.0	Microbicide, worse than placebo, is preventive compared with "condom only"
H	RR = 1.0	RR < 1.0	Microbicide is equivalent to placebo, and both prevent HIV infection compared with "condom only"
I	RR = 1.0	RR > 1.0	Microbicide is equivalent to placebo, and both are worse than "condom only"

\* HIV incidence compared in experimental microbicide arm vs control arm: "RR < 1.0" denotes increased risk in the control arm as compared with the experimental microbicide arm; "RR > 1.0" denotes decreased risk in the control arm compared with the experimental arm; and "RR = 1.0" denotes approximately equal risks of infection in the experimental and control arms.

† These interpretations assume a) that true levels of condom use do not vary across trial arms, and (b) that self-reports of condom use reflect reality. With a necessarily unblinded "condom-only" arm, both true levels and self-reports of their use might well vary systematically across trial arms, rendering these comparisons uninterpretable (see text).

devastation of the HIV epidemic justifies pressing forward as quickly as possible to a judgment about each of the proposed test substances. Sponsored and funded by different institutions, the trials emanate from different countries and operate at different sites. Given this multiplication, the issue of concern is how best to achieve the common goal of a proven and safe microbicide.

At the time of this writing, we anticipate there will be at least five trials, each testing different substances. We have no way at this stage to choose among the five or six substances to be tested. Several of the test microbicides are rather similar in supposed modes of action or differ only in dosage. To distinguish among them, however, demands the expensive and laborious task of conducting field trials.

Each trial will require a large population to be tested. Prophylactic trials must assemble an unaffected population in which the numerator will be the cases not prevented. The large numbers required are a difficulty alleviated only by very high incidence rates or by highly effective prevention. There are few populations in which enough incident HIV cases can be expected to arise at any one site over a reasonably brief observation period. To achieve the required numbers, several of the individual trials therefore incorporate three or four different sites, each using similar designs.

Given this necessity for several sites, statistical adjustments for differences across sites will of course be required. Yet adjustments are unlikely to assure complete control of factors that may yet produce differences between sites. Hence, in a trial with several sites the populations should be as homogeneous as possible. The same principle holds for assessing the performance of different substances. If their effects are to be compared across different trials, homogeneity among trials in all attainable ways is desirable.

In practice, neither of these economies is achievable in full. The difficulties of execution, the variation in

expected incidence rates, and the large administrative and funding issues preclude the ideal situation. This is not to say that variations in circumstance and procedures can yield no advantages. The issues are well stated in John Stuart Mill's "canons," logical strategies he devised for inferring causation.<sup>7,8</sup>

Two of the canons are relevant here. His second canon, the "method of difference," can be briefly restated: the situations compared are alike in all variables but one. This is the rigorous standard to which randomized controlled trials aspire and for which we have argued above. Mill's first canon, here briefly restated, is the "method of agreement." The situations compared are consistent in having only one circumstance in common. In other words, if a putative microbicide has the same effect across different situations, the result offers grounds for a causal attribution. This canon is a saving grace for causal inference in observational epidemiology. Given adequate rigor, even in different situations replication of a result adds strength to a causal attribution.

Nevertheless, in randomized controlled trials, the more similar the testing procedures are to each other, the more it will be possible to consider the results of the different trials as reinforcing the results of each, whether confirmatory or rejecting. In this light, the choice of placebo becomes a key issue in comparisons across trials. Every effort should be made to employ an identical substance as placebo.

For each trial, the placebo should look, smell, taste and be of the same viscosity as each putative microbicide. Admittedly, the degree to which the vehicles for a test product differ across trials could make it difficult to achieve complete uniformity among placebos. This does not relieve the trialists from making the effort to strive for uniformity. To this end, the composition of each placebo should be knowledge freely shared across trials.

In practice, too, several of the current trials deliver the test substance and placebo in applicators, which confers some flexibility in the degree to which the appearance of each needs to be identical. Given the unplanned but welcome contemporaneous introduction in current trials of several putative microbicides, there is every reason to provide to the extent possible a standard that would permit direct comparisons of effect.

These trials are critical in dealing with the worldwide contagion of HIV/AIDS. The need is pressing to demonstrate the effectiveness of interventions, for microbicides as for vaccines. Epidemiologists have a duty to produce sound study designs. Statisticians are an invaluable and indispensable resource in such endeavors. They can advise on and perhaps perform the most appropriate analytic techniques, as well as those statistical refinements (multiple adjustments, metaanalyses, etc.) that help to correct for bias and the like. Nevertheless, epidemiologists have to know that designs flawed in the first place will inevitably weaken inference and interpretation. Epidemi-

ology lives or dies by the rigor, integrity, economy and good judgment applied to research design.

## References

1. Stein Z. HIV prevention: the need for methods women can use. *Am J Public Health* 1990;80:460-462.
2. The Microbicide Initiative. *Mobilization for Microbicides: The Decisive Decade*. New York: The Rockefeller Foundation, 2002.
3. Roddy RE, Zekeng L, Ryan KA, Tamoufe U, Tweedy KG. Effect of nonoxynol-9 gel on urogenital gonorrhea and chlamydial infection: a randomized controlled trial. *JAMA* 2002;287:1117-1122.
4. Richardson EA. Nonoxynol-9 as a vaginal microbicide for prevention of sexually transmitted infections: it's time to move on. *JAMA* 2002;287:1171-1172.
5. Van Damme L, Ramjee G, Alary M, et al. Effectiveness of COL-1492, a nonoxynol-9 vaginal gel, on HIV-1 transmission in female sex workers: a randomised controlled trial. *Lancet* 2002;360:971-977.
6. WHO/CONRAD. *Technical Consultation on Nonoxynol-9. Summary Report*. Geneva: World Health Organization, 2001.
7. Mill JS. *A System of Logic*. London: Parker, Son and Bewin, 1856.
8. Susser M. What is a cause and how do we know one? A grammar for pragmatic epidemiology. *Am J Epidemiol* 1991;133:635-648.

# Commentary: The Design of Prophylactic Trials for HIV

Nancy Padian

In this issue of EPIDEMIOLOGY, Zena Stein and her colleagues<sup>1</sup> raise questions about the current plans to include two control groups in randomized clinical trials (RCTs) of topical microbicides. (Microbicides are topical vaginal applications intended to protect women from HIV infection.)

Clearly, the results of an RCT are more difficult to interpret when there are three rather than two treatment arms. Multiple arms are also more expensive and challenging to implement uniformly across multiple sites. Given scarce resources and the urgency of finding a microbicide that effectively prevents the transmission of HIV, the authors make a compelling argument for a

two-arm trial. Their recommendation is to omit the condom-only arm and to make the head-on comparison of an experimental microbicide product and a placebo. The simplicity of such a trial might appear preferable. However, a similar argument could be made to eliminate the placebo arm and simply compare the experimental product with condoms only.

Without a condom-only arm, it is more likely that a product providing more protection than condoms could be eliminated from consideration unless the product were also more effective than a placebo (see the authors' Table 1, specifically row G).<sup>1</sup> Similarly, if a product were as effective as condoms, even if it were less effective than a placebo (row F), it could provide a viable female-controlled method of prevention.

Perhaps even more critical is the interpretation of row E in the table, which implies that use of only the placebo arm would result in promotion of a product that is less effective than condoms. (However, it is difficult to imagine this scenario given that women in all study arms

---

From the Department of Obstetrics, Gynecology and Reproductive Sciences, University of California, San Francisco, San Francisco, CA.

Address correspondence to: Nancy Padian, Department of Obstetrics, Gynecology and Reproductive Sciences, University of California, San Francisco, 74 New Montgomery Street, Suite 600, San Francisco, CA 94105-3444; NPadian@psg.ucsf.edu

are counseled to use condoms, thus allowing only the evaluation of the marginal effect of a product beyond condom use.)

It is true that the "open nature" of a condom-only arm suggests that "outcomes could very well reflect differences in behavior rather than in any one of the interventions tested."<sup>1</sup> Because a comparison with a condom-only arm cannot be blinded on any level, women in the microbicide arm will know they are in an experimental intervention. They may be more likely to reduce their risky behavior than women in the condom-only arm who are not receiving a new intervention. However, as long as there is some measure of protection, does it really matter whether it is attributable to behavioral change or to the product itself?

Initially, it seems more important to know that the intervention works than to be able to identify the particular aspect of the intervention that results in a beneficial effect. If a microbicide were to be found more effective than condoms, then future studies (by comparing this effective product with a placebo) could consider whether the benefit was attributable to the microbicide itself or to concurrent behavioral changes.

The correct comparison group should be directly linked to the primary aim of study. Comparisons with a condom-only arm address the research objective of finding something that is better than (or at least as good as) the only known effective method to prevent heterosexual transmission of HIV. Comparisons with a placebo arm primarily address the research objective of whether

an active agent is more effective than an inert placebo, controlling for behavior. A three-arm trial allows both issues to be addressed simultaneously. Ultimately, for example, the question of whether a placebo is really inert (the "design hazard" noted for the Col-1492 study)<sup>1</sup> might best be addressed by comparing the placebo with condoms only. Had this comparison been possible, and if the placebo were protective, we might currently have one effective product, whereas now we have none.

All this said, if only two arms are possible, there is a persuasive reason why use of a condom-only arm might be problematic. Women in a condom-only arm may not feel that they are really getting any kind of intervention, and therefore they may be less likely to participate in a long-term trial. This could cause differential dropout and follow-up rates that could ultimately invalidate the original random assignment and statistical comparisons of the study's results. This clearly merits further study.

---

### About the Author

NANCY PADIAN is a professor in the Department of Obstetrics, Gynecology and Reproductive Sciences and director of international programs at the AIDS Research Institute, University of California-San Francisco. For almost 20 years, she has developed and directed national and international projects on reproductive health in young women, most currently the effectiveness of barrier contraceptives and microbicide products to slow the heterosexual transmission of HIV and other STDs.

---

### Reference

1. Stein Z, Myer L, Susser M. The design of prophylactic clinical trials: The case of microbicides for HIV. *Epidemiology* 2003;14:80-83.