

# FDA Antiviral Drugs Advisory Committee

20 August 2003

TMM Farley  
Department of Reproductive Health and Research  
World Health Organization  
1211 Geneva 27  
Switzerland

---

The World Health Organization is involved in research on the assessment of the safety and effectiveness of microbicides in developing country populations. This work is being conducted in partnership with the CONRAD Program, Arlington, VA, with primary support from the Joint United Nations Program on HIV/AIDS (UNAIDS) and the United States Agency for International Development (USAID).

I would like to address three specific issues: measures of product effect, choice of control arms, and strength of evidence necessary to demonstrate effectiveness.

## Measures of product effect

### *Efficacy*

If data were available on the incidence of HIV infections in the study cohort when condoms are used, then a simple one-arm observational study in which the active product is used in addition to condoms would provide the necessary information on **efficacy**, which is defined as the proportion of infections prevented. However, such data are seldom available and thus a concurrent control group is necessary to estimate the incidence of infections when condoms alone are used.

In practice, since volunteers have different inherent risks of infection which depend on their personal characteristics, their sexual behaviour and that of their partners, randomization would be used to ensure that high-risk and low-risk volunteers are balanced across the two study groups. If the effect of the product were to reduce the risk of infection by the same amount (e.g. two-fold reduction in risk) among high-risk and low-risk volunteers, then a comparison of the infection rates in the product and no product arms would also provide an estimate of efficacy.

In addition, condoms are not always used or always used correctly and they can slip or break. If the effect of the product is the same under these different conditions, then the trial can estimate the efficacy of the product directly, provided the proportions of acts of intercourse when condoms are not used, incorrectly used, or slipped or broke are balanced between the study groups. Randomization can ensure balance in the proportions of volunteers likely to have problems with condom use, or likely not to use the condoms consistently.

### *Effectiveness*

Product efficacy is defined as the proportion of infections prevented when the product is used correctly. In practice the product may not be used for all acts of intercourse, or may not be used

correctly (e.g. not inserted at the appropriate time interval before exposure, or not inserted sufficiently deeply into the vagina so that some product leaks). In these circumstances, it is not possible to estimate the efficacy of the product, but only to estimate its **effectiveness**. This is defined as the proportion of infections prevented when the product is used in a typical fashion, which is a mixture between perfect use and imperfect use. The effectiveness of the product is always less than its efficacy.

### ***Use-effectiveness***

An additional measure of the impact of a microbicide product is its **use-effectiveness**. This is the proportion of infections prevented when the product is used beyond the carefully controlled conditions in a research study, in a more general population of users. Use-effectiveness could include such factors as: (a) impact of incorrect product use by women who are less well trained or informed as study volunteers, (b) changes in behaviour due to less consistent condom use, or (c) inconsistent supply of the product. In general, use-effectiveness will be less than effectiveness. In addition, use-effectiveness is very difficult to estimate and assess since directly comparable information are lacking, and attempts to collect information in a systematic manner is likely to distort behaviours.

### **Choice of control arm**

Since it is well known that condoms reduce the risk of HIV infection, it is not ethical to conduct a trial of a new microbicide without strongly promoting condoms as well as instructing volunteers on how to use them correctly. Thus an ethical microbicide trial can at best estimate the effect of using the product as an addition to using condoms.

### ***Masking or blinding***

A key problem in assessing the efficacy or effectiveness of microbicides is the complexity and interplay of factors which determine whether the product is used, or used correctly, and whether or not a condom is used, or used correctly, for each act of intercourse. There is evidence from studies of condoms that they are more likely to be used when there is a perceived high risk of infection, and less likely to be used when the perceived risk is low. With the addition of a microbicide, there are more possibilities – condom and product together, condom alone, product alone, or neither. While randomization can ensure that volunteers likely to not use condoms, or likely not to use condoms correctly, are balanced between the study groups, it cannot ensure that the likelihood of condom use remains the same after randomization. There may be interactions between product and condom use. For example, some users may consistently use both condoms and product together, or neither on the rare occasions when they do not have access to supplies. Other users may be more likely to use the product or condoms, but seldom use them both together. It is very difficult to predict *a priori* which pattern of use may happen. The only way to ensure balance in post-randomization characteristics and behaviours is to ensure that the comparison or control group receives a **placebo** product. Under ideal circumstances the placebo product cannot be distinguished from the active product and the user remains **masked or blinded** as to which product is used. In such circumstances the study will be able to provide an unbiased estimate of the effectiveness of the product (proportion of infections prevented in the active compared with the placebo arm).

**Recommendation:** If a placebo product exists and masking can be maintained throughout the study, it is preferable to conduct a randomized double-blind trial. This gives an unbiased estimate of product effectiveness, and randomization ensures that factors which may influence

infection rates are balanced. Masking ensures that post-randomization factors remain balanced also.

### ***Information on condom and product use***

In order to estimate effectiveness, it is not strictly necessary to collect information in a double-blind trial on the actual patterns of product use and condom use. By assumption these are balanced across the two study groups. In practice, it would be wise to have measures of compliance with product use and patterns of condom use, if only to check that these are indeed balanced. If not, then some adjustments will have to be attempted, but it must be recognised that these are at best partial adjustments and that the observed effectiveness will be lower than the true effectiveness due to misclassification and potential misreporting of condom and product use. It is not possible to know the extent of misclassification and its final impact on the estimated effectiveness, but it is very unlikely that complete and accurate behavioural information can be guaranteed.

If reported product and condom use are similar between the study groups, then the study can provide a direct unbiased estimate of effectiveness. It is not necessary for the primary estimate of effectiveness to use the behavioural information, but this would form the basis of supplementary analyses to assess internal consistency, or identify particularly interesting subgroups.

If a placebo product is not available, a double-blind trial cannot be implemented and there is no alternative to a parallel control group which uses condoms only. Randomization will ensure balance of baseline characteristics, but cannot guarantee balanced behaviours or reporting once the study allocation has been revealed. In this circumstance the primary study analysis and estimate of effectiveness must make whatever adjustments are possible for the reported compliance with product use and the patterns of condom use. The validity of the analysis will be unknown, and the inferences less compelling. It is much more important in a study which cannot be double blind to ensure that the behavioural data are complete and accurate, and the study team must be able to demonstrate the validity of these data. In a double-blind study, validating the behavioural information is desirable though not essential.

### ***Number of control groups***

The discussion above argues strongly for using a placebo product to maintain masking throughout the trial whenever such a product is available. The analysis and interpretation of the data are much more straightforward and give rise to an unbiased estimate of product effectiveness.

If a placebo product is available, there is nothing to be gained by including a no product or condom-only arm in the trial. Not only will this add cost and complexity to the trial and delay the assessment of effectiveness, but the analysis will be complex and confusing. For the comparison between the active and placebo products, the intent-to-treat or 'as randomized' analysis is the most compelling. The adjustment for the behavioural data is of less importance. By contrast, for comparisons between active and no product, or placebo and no product, the primary analysis would require adjustment for reported behaviours.

Even if the true incidence rates in the placebo and no-product groups were similar and higher than the incidence in the active product group, the final analysis may demonstrate a significant difference between the active and placebo groups, but no difference between the active and no

product groups. This would be expected as a result of misclassification of reported behaviours which will dilute the estimated differences. The comparison between the active and no product groups would not make any useful contribution to understanding the difference between the active and placebo groups.

**Recommendation:** If a placebo product exists such that a double-blind trial can be conducted, there is no value to including an additional no-product control group. If no placebo product exists, there is no alternative to using a no-product control group.

### **Example**

The COL-1492 trial<sup>1</sup> compared a gel containing the spermicide N-9 with a similar, non-spermicide gel. The study showed a significant difference between the two study groups. The analysis of the behavioural data was complex and demonstrated internal consistency in the patterns of risk differences. As a result of the study, users are advised against N-9 for preventing HIV. Inclusion of a third, no-product arm in the trial would have added considerably to the complexity of the trial and the inferences. The result from a no-product arm may have informed whether the placebo product appeared to have any beneficial effect, but would not have changed the interpretation of the primary study result that N-9 was associated with a significantly higher incidence of HIV infection than placebo.

### **Strength of Evidence**

The need for unequivocal and convincing data on the effectiveness of a new microbicide must be balanced against the public health imperative to make promising products rapidly available to women at risk of HIV infection. The HIV pandemic has forced priorities to be reconsidered and the FDA must be commended for its flexibility and willingness to consider the imperatives of the disease.

Under normal circumstances, the FDA has stated that it requires two independent trials demonstrating effectiveness at the  $\alpha = 0.05$  level. Since it may not be ethically possible to conduct a second trial once a first well-conducted trial has demonstrated effectiveness at the 0.05 level, the two trials must either be conducted simultaneously, or a single trial must by itself provide more convincing evidence.

Two independent trials each significant at the  $\alpha = 0.05$  level correspond to an overall p-value of 0.0013 ( $= 2 \times [0.05/2]^2$ ). This is a very stringent requirement for demonstrating significance, far beyond conventional levels. A well-conducted study, with good internal consistency with such a final p-value would be very compelling. But, would a less stringent p-value be sufficiently compelling? Not only would this save scarce resources for product testing, but it would shorten the time to make a new product available.

The purpose of the microbicide is to prevent HIV infection. Phase 3 studies necessarily use HIV infection as the primary endpoint. There are no recognised surrogate endpoints. At present HIV infection is not curable and volunteers infected during a study have little chance of accessing care. If a study is going to demonstrate a significant difference between groups, then

---

<sup>1</sup> Van Damme L, Ramjee G, Alary M, *et al.* Effectiveness of COL-1492, a nonoxynol-9 vaginal gel, on HIV-transmission among female sex workers. *Lancet* 2002;**360**:971-977.

half the volunteers receive the inferior product. It is not ethical or acceptable to expose volunteers to a high risk of infection more than necessary. Data and Safety Monitoring Boards regularly grapple with the difficult balance between allowing trials to continue to a scientifically convincing conclusion and the need to stop exposing volunteers to an inferior product as rapidly as possible. It is never an easy decision.

Assessing the balance of risks and benefits of research is also the responsibility of Ethical Review Committees (ERCs) which must approve the research protocol before implementation and receive regular reports on progress. It is likely that institutional ERCs will query whether a p-value of 0.0013 is strictly necessary, particularly for prevention of a fatal disease and no other products are available. If it is considered unethical to allow a second trial when a first study has demonstrated effectiveness at the  $p = 0.05$  level, then it is equally unethical to require a single trial to provide evidence to the strength of two independent trials.

The example of the COL-1492 trial may be helpful here. The study showed a higher risk of infection in the N-9 group compared with placebo with a result that was just significant at the  $p = 0.05$  level. At the conclusion of the trial, some commentators were not convinced that N-9 was definitely harmful, suggesting that a single p-value at the 0.05 level does not carry sufficient weight. In this circumstance, there were other data about the effects of N-9 on the risk of HIV infection which helped interpret the results. With a new microbicide there may be no supporting data.

It is interesting to speculate what may have happened if the COL-1492 trial result had been reversed but with the same p-value. There would have been strong pressure to make the product rapidly available to potential users, but it is also likely that residual concerns about the strength of evidence would remain. This suggests that evidence to the level  $p = 0.05$  is not adequate by itself. It does not however tell us how small a p-value must be in order to be convincing.

**Recommendation:** The Advisory Committee is strongly urged to consider the ethics of requiring evidence to the level of  $p = 0.0013$ . A conventional single trial p-value of 0.05 may not be adequate by itself, but it is unclear how small a p-value must be. It can be argued that the proposed 'compromise' p-value of 0.01 is also ethically questionable, even though it is 10-fold higher. Since the conventional p-value of 0.05 is arbitrary, it is difficult to determine a truly objective level. The FDA is urged to be flexible. Results must be interpreted carefully in the context of all available information.