

Advocates' Guide to Statistical Terms

No study can produce a simple “yes” or “no” on whether or not a product “worked”. To decipher the headlines and discussions regarding the data from this or any trial, it is useful to understand some statistical terms used to describe a trial result. For the MDP 301 study, the data analysis will include comparisons of rates of infections in participants enrolled in the PRO 2000 and placebo groups.

One key term is **statistical significance**. If a result is described as statistically significant, it means that an observed difference (for example, between two arms of a microbicide trial) is very likely due to the product itself and is not a coincidence. Significance is always given with a **confidence level**. A 95 percent confidence level, which is standard for many trials, means that there is at most a 5 percent likelihood of a statistically significant result having occurred solely by chance.

The trial team will also report on the **confidence intervals** associated with its findings. A confidence interval is a way of describing the reliability of the finding, which is given as a **point estimate** – such as a 35 percent reduction in risk of infection. The narrower the confidence interval around the point estimate, the more likely it is that the result is accurate and would be seen again if the trial was repeated.

This can be confusing because all these values are inter-related, but to fully understand the strength of a result, one must know: (1) the point value; (2) whether or not the result is statistically significant; (3) the confidence level, which may be expressed as a percent (95% or more) or a p-value (.05 or less); and (4) the confidence interval.

Other terms that might be used in discussion of initial results are **per-protocol** (PP) and **intent-to-treat** (ITT) analyses of results. Per-protocol results only include infections that occurred in those participants who received the full intervention. With a user-controlled product, where use in a trial is measured through self-reporting, it is very difficult if not impossible to determine a per-protocol result. Intent-to-treat counts every infection after trial enrollment, regardless of whether a participant used the product, the placebo, or condoms. ITT is considered a gold standard because it considers every randomized participant. ITT analyses remove the risk that dropouts or subgroups may have unbalanced the two arms to any degree. In general, the safest route is to report both PP and ITT and to analyze any difference.

In a microbicide trial, an approximation of a “per protocol” result would be a sub-group analysis of participants who reported high gel use and low condom use. This is controversial in microbicide trials for several reasons: it is not known whether this group differs in other ways from the rest of the trial population (thereby undermining randomization); it is not clear how accurate the self reporting of gel and condom use is; and the numbers in any sub-group analysis are unlikely to be large enough to allow for a statistically significant result.

Advocates' take-home message: no matter what the headlines say, a single number is not the full result.