35. Statistical significance vs. clinical significance

François Noël, June 2026
(Reviewed by José Eduardo Tanus dos Santos, MRP, USP)

In pharmacology, it is very common to compare a control group and an experimental group to determine whether a drug has induced a response, using statistical methods to assess whether this response occurred merely by chance or not (i.e., whether it is "statistically significant"). If the data follow a normal (Gaussian) distribution, statistical analysis is generally performed using Student's *t*-test or a similar method, generating a *P* value used to identify an effect (Curtis *et al*., 2015).

> Historical note: The name of the test comes from the pseudonym "Student", used by William Sealy Gosset, a statistician who worked at the Guinness brewery in the early twentieth century. He signed the paper *"The probable error of a mean"* (*Biometrika* 6(1):1–25, 1908) as Student because the company did not allow employees to publish scientific findings under their own names.

Statistically significant difference

Far too often, however, the *P* value is reduced to a dichotomy in which results are declared "statistically significant" when *P* is equal to or below a cut-off point (generally 0.05 in the biomedical sciences) and "non-significant" otherwise, a practice that has become particularly pernicious in statistical analysis (Amrhein *et al*., 2019; Greenland *et al*., 2016). For this reason, it is recommended that exact *P* values be reported, rather than merely indicating whether they are above or below 0.05 (or another threshold), thereby allowing readers to interpret values close to the cut-off appropriately, such as 0.049 and 0.051.

Moreover, the use of the cut-off point $P = 0.05$, traditionally accepted in the biomedical sciences, is now recognised as largely conventional, although widely criticised. Authors such as Colquhoun go (much) further by advocating the adoption of $P < 0.001$ in order to substantially reduce the risk of false-positive findings (*apud* Curtis *et al*., 2015).

Rather than focusing exclusively on the *P* value, several authors strongly recommend reporting the effect size together with its corresponding confidence intervals (see below).

Clinically significant difference

Even when statistical analyses indicate a difference, this does not necessarily imply that it is clinically significant, a concept whose definition is far from straightforward. Indeed,

in contrast to the relatively well-established standards (albeit debatable, as discussed above) used for decisions regarding statistical significance, there are no specific guidelines to determine what magnitude of difference should be considered "clinically significant" or "practically important" (Kaul and Diamond, 2010). Thus, the threshold for clinical relevance depends on the context: for example, a 10% variation may be highly relevant in certain gene expression studies, whereas a 90% reduction in viral titre may still be irrelevant in some viral infection studies (Curtis *et al*., 2015).

Once the **Minimum Clinically Important Difference (MCID)** has been estimated for a given treatment, the confidence interval of the relative risk (risk ratio: RR) may be used to assess whether the treatment provides both statistically and clinically significant benefit (Fig. 1).

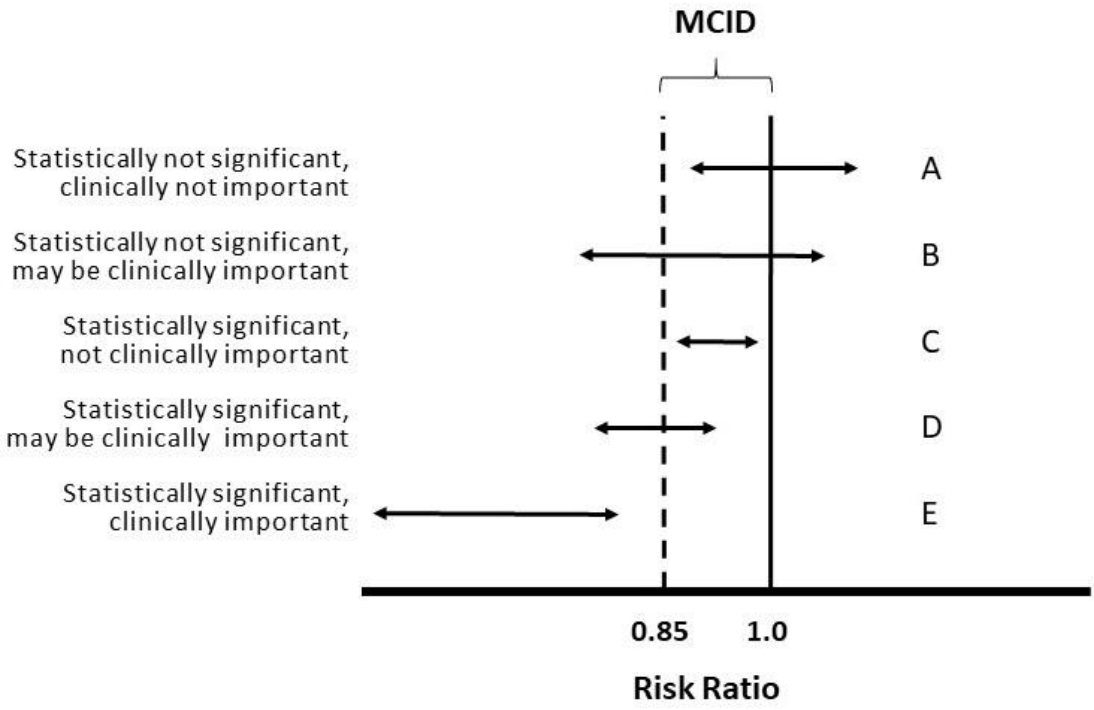

Figure 1. Graphical representation of therapeutic benefits that are statistically significant and clinically important. Five trial outcomes (A-E) and their interpretation in relation to the null effect (risk ratio = 1.0) and the Minimum Clinically Important Difference (MCID) are presented, corresponding to a 15% relative risk reduction (equivalent to a risk ratio of 0.85). Treatment effects (double-headed arrows) are expressed as 95% confidence intervals. *(*Adapted from Kaul and Diamond, 2010*).*

Another widely used index for assessing the clinical importance of a treatment is the Number Needed to Treat (NNT). In general, therapeutic interventions are considered clinically important when the NNT is < 50 (Kaul and Diamond, 2010).

References

Amrhein V., Greenland S., McShane B. Scientists rise up against statistical significance. *Nature* 567(7748):305–307, 2019.

Curtis M.J., Bond R.A., Spina D., Ahluwalia A., Alexander S.P.A., Giembycz M.A., *et al.* Experimental design and analysis and their reporting: new guidance for publication in BJP. *Br. J. Pharmacol.* 172:3461–3471, 2015.

Greenland S., Senn S.J., Rothman K.J., Carlin J.B., Poole C., Goodman S.N., *et al.* Statistical tests, *P* values, confidence intervals, and power: a guide to misinterpretations. *Eur. J. Epidemiol.* 31:337–350, 2016.

Kaul S., Diamond G.A. Trial and error: how to avoid commonly encountered limitations of published clinical trials. *J. Am. Coll. Cardiol.* 55(5):415–427, 2010.