

Comment on *Probabilities of Clinical or Practical Significance*

Alan M Batterham

Sportscience 6, sportsci.org/jour/0201/amb.htm, 2002 (725 words)

Department of Sport and Exercise Science, University of Bath, Bath BA2 7AY, United Kingdom. [Email](#).

[Reprint pdf](#) · [Reprint doc](#)

Will Hopkins' [article](#) provides a timely and valuable elaboration of the [short item](#) in the previous edition of Sportscience. Up to now, forward thinking researchers have been able to access a spreadsheet at [newstats.org](#) to calculate likelihoods of an observed effect being clinically or practically beneficial, trivial, or harmful (derived from the t distribution). This novel approach has extended our battery of tools for drawing inference from data, beyond both the outmoded tests against the null hypothesis and more contemporary estimation methods using confidence intervals or limits. The current article, with the associated table, provides a tool for illuminating the probabilities derived from the spreadsheet. The qualitative descriptors are an important enhancement, facilitating interpretive statements in the discussion and conclusions sections of the dissemination of research findings. I strongly encourage readers attempting to get to grips with these concepts to read this article alongside the previous short item, and to view the PowerPoint slide show. The benefits of a research design, data analysis, and interpretation approach based on quantifying clinical significance, rather than mere statistical significance, have been recognized widely in the medical sciences. Unfortunately, the message has not been widely adopted in the exercise science field. This article will hopefully encourage more researchers in our field to adopt this approach, and challenge some cherished assumptions and dogma. It will require a concerted effort by many people to bring about this paradigm shift away from the comfort zone of $P < 0.05$.

The summary of advice provides a clear framework for reporting of research findings. The fourth bullet point often presents the biggest challenge for researchers and, indeed, clinicians and practitioners. Determining, in advance of the study, the minimum clinically important difference (MCID), or smallest worthwhile effect, is not a trivial issue. However, it is one that must be tackled if any true insight is to be derived from the research. Firstly, knowledge of the MCID, combined with a value for the 'noise' in the measurement from a reliability study, permits the calculation of an appropriate sample size for adequate precision of estimation. Secondly, knowledge of the MCID allows for the calculation of the probabilities of clinically important/ trivial/ or clinically harmful effect in the population, with the associated qualitative descriptors from the table.

I am particularly moved by the last paragraph regarding the choice of an appropriate confidence interval to convey precision. The most commonly used are the 95% and 99% confidence intervals. These values are arbitrary and have been widely adopted, primarily, due to their congruence with null hypothesis testing at the 0.05 or 0.01 alpha levels. In other words, if the 95% or 99% confidence interval does not contain the value assumed under the null or zero, then P is < 0.05 or < 0.01 , respectively. In my view this practice should be strongly discouraged and an alternative philosophy adopted. I agree that the 95% limits are too high and often give a false impression of imprecision. The presentation of 50% likely limits, or "possible" limits for the true effect represents a radical, yet welcome, departure from conventional practice. A confidence interval is defined by the probability that it contains the true or population value. Hence, a 50% CI is the interval that you are 50% certain contains the value that would be estimated from a much larger study. Confidence of only 50% may be an anathema to those locked into testing null hypotheses, as, in their eyes, it is equivalent to testing against the null

hypothesis at a P value of 0.5. However, as stated, this alternative philosophy should not be viewed as an analogue of significance testing.

In his concluding sentence, the author doubts whether 50% likely limits will come into widespread or, indeed, any use during his lifetime. As with the adoption of analysis based on clinical significance, how fast and how far we progress is in our own hands. Active engagement with the gatekeepers of knowledge—the reviewers and journal editors—together with continuing education of ourselves, our faculty peers, and our students, may help shift the paradigm.

[Back to article/homepage](#)

Published July 2002
[editor](#)
[©2002](#)