Magnitude-based inference: What is it? How does it work and is it appropriate?

B. Van Hooren [1]

[1]NUTRIM School of Nutrition and Translational Research in Metabolism, Maastricht University Medical Centre+, Department of Nutrition and Movement Sciences, Maastricht, The Netherlands

Magnitude-based inference | Video | Tutorial

Headline

Research in the field of sports science is frequently performed on a relatively small number of individuals. We are usually however interested in knowing whether the effect found in our sample of individuals also applies to a larger group: the population from which the sample is drawn. For this purpose, we use statistical inferential methods. There are several statistical inferential methods available. The most widely used method is arguably null-hypothesis significance testing. This method has been widely criticized since its introduction, most prominently because statistically significant results are not necessarily clinically relevant and statistically non-significant results can still be clinically relevant (Figure 1). (1, 2)

Magnitude-Based Inference. Motivated by the limitations of null-hypothesis significance testing, Batterham and Hopkins (1) developed a new statistical inferential method in 2006 entitled *"Magnitude-Based Inference"*. In this method, confidence intervals are interpreted in relation to a smallest worthwhile change (Figure 1). The method has seen a large uptake in the sports science community and is also increasingly used in other research fields. Despite this large uptake, not all researchers and practitioners fully understand how the method works. Understanding how Magnitude-Based Inference works is however important as it helps researchers and practitioners to correctly interpret the results of studies that have used this method. In a new video, I therefore explain what Magnitude-Based Inference is and how it works. Click here for the link to the video.

Criticism

Several researchers have criticized Magnitude-Based Inference, in particular for interpreting a frequentist confidence interval as a Bayesian credible interval and for having high rates of type I errors (false positive were you conclude there is a substantial effect while there is no substantial effect).(3-6) These researchers therefore advised to use other statistical inferential methods such as a full Bayesian analysis (which has recently been performed with Magnitude-Based Inference (7)) or equivalence testing (see for example (8)). Batterham and Hopkins have responded to these criticisms. (9-13) They justify their Bayesian inferences with the confidence interval by claiming to use Bayesian methods, but with a non-informative prior distribution which results in the Bayesian credible interval to be equivalent to the frequentist confidence interval, when several other assumptions are met.(5) Further, they re-analysed the inferential error rates using different definitions of the errors (12) as used by the other researchers (4) and argued that the definitions of errors used in a recent critique paper (6) are also not entirely appropriate.(13)

The debate around these statistical issues may be difficult to follow for a sport scientist who does not have a strong background in statistics. Understanding this debate is however important as it allows researchers to decide on whether they should use Magnitude-Based Inference or other statistical inferential methods for their studies. Therefore, in a second video I discuss some of the criticisms on Magnitude-Based Inference and the responses by Batterham and Hopkins in a (hopefully) understandable way. Click here for the link to the second video.

Other approaches

Researchers that like the idea of Magnitude-Based Inference, but who do not want to use it based on the criticisms can use several other methods which are roughly similar. For example, Mengersen, Drovandi, Robert, Pyne and Gore (7) recently performed a full Bayesian analysis of Magnitude-Based Inference with a flat prior distribution. Other approaches that are roughly similar to Magnitude-Based Inference include using regions of practical equivalence (ROPE) in a Bayesian approach or equivalent testing in a frequentists approach (8, 14).

Twitter: Follow Bas Van Hooren @BasVanHooren

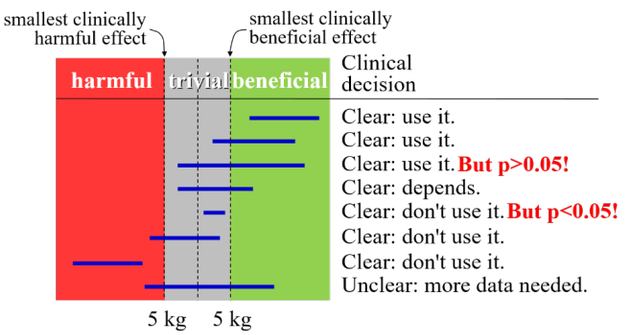

Fig. 1. Magnitude-Based Inference. Decisions in Magnitude-Based Inference are made based on confidence intervals (represented by the blue horizontal lines) in relation to a smallest worthwhile change (represented by the dashed vertical lines on each side of the trivial area). Consider the following example: a study has investigated the effects of 4 weeks resistance training on back squat 1 repetition maximum performance. Any increase or decrease larger than 5 kg is considered relevant, while all changes smaller than 5 kg are too small to be of practical relevance (i.e., trivial). In Magnitude-Based Inference, confidence intervals define the likely range of the population value. If the study finds the effect illustrated by the first confidence interval, the conclusion is therefore that the intervention is (very likely) effective as the confidence interval is entirely in the beneficial area. For the second interval, the confidence interval overlaps the trivial and beneficial areas. However, the overlap in the beneficial area is larger and the intervention is therefore more likely to be beneficial as trivial. The conclusion could therefore be to use the intervention because it might be beneficial and in the worst case scenario the intervention will have a trivial effect. In the 3rd and 4th interval, the overlap of the confidence interval into the trivial area has increased, so the intervention could have a trivial effect, but it could also be beneficial. If the training intervention would not require much time and money, a coach could still decide to use the intervention. However, the results are not statistically significant. Conversely, in the 5th interval the intervention has likely only a trivial effect, but it is significant. These latter examples illustrate the mismatch between practical relevance and statistical significance. Figure adapted from Batterham and Hopkins (1).

References

1. Batterham AM, Hopkins WG. Making meaningful inferences about magnitudes. Int J Sports Physiol Perform 2006;1(1):50-57.
2. Stang A, Poole C, Kuss O. The ongoing tyranny of statistical significance testing in biomedical research. Eur J Epidemiol 2010;25(4):225-30.
3. Barker RJ, Schofield MR. Inference about magnitudes of effects. Int J Sports Physiol Perform 2008;3(4):547-57.
4. Welsh AH, Knight EJ. "Magnitude-based inference": a statistical review. Med Sci Sports Exerc 2015;47(4):874-84.
5. Butson M. Will the numbers really love you back: Re-examining Magnitude-based Inference. 2017
6. Sainani KL. The Problem with "Magnitude-Based Inference". Med Sci Sports Exerc 2018
7. Mengersen KL, Drovandi CC, Robert CP, et al. Bayesian Estimation of Small Effects in Exercise and Sports Science. PLoS One 2016;11(4):e0147311.
8. Lakens D. Equivalence Tests: A Practical Primer for t Tests, Correlations, and Meta-Analyses. Soc Psychol Personal Sci 2017;8(4):355-62.
9. Hopkins WG, Batterham AM. An imaginary Bayesian monster. Int J Sports Physiol Perform 2008;3(4):411-2.
10. Batterham AM, Hopkins WG. The case for magnitude-based inference. Med Sci Sports Exerc 2015;47(4):885.
11. Hopkins WG, Batterham AM. Magnitude-based inference under attack 2014 [Available from: sportsci.org accessed 28-2-15.
12. Hopkins WG, Batterham AM. Error Rates, Decisive Outcomes and Publication Bias with Several Inferential Methods. Sports Med 2016;46(10):1563-73.
13. Hopkins WG, Batterham AM. The Vindication of Magnitude-Based Inference (draft 2) 2018 [Available from: sportsci.org/2018/mbivind.htm accessed 16-5- 2018.
14. Kruschke JK. Rejecting or Accepting Parameter Values in Bayesian Estimation. Advances in Methods and Practices in Psychological Science: 251524591877130, 2018.