

- [Exercise Science at This Site](#) with a new associate editor.
- [Magnitude-Based Decisions](#): rebranding MBI; dismissing another critique.
- [Reprint pdf](#) · [Reprint docx](#)

Exercise Science at This Site

Ross D Neville, School of Public Health, Physiotherapy and Sports Science, University College Dublin, Dublin, Ireland. [Email](#). Reviewer: Will G Hopkins, Institute for Health and Sport, Victoria University, Melbourne, Australia. Sportsscience 23, i, 2019 (sportsci.org/2019/inbrief.htm#exsci. Published June 2019. ©2019

The Sportsscience journal site welcomes Dr Ross Neville, as associate editor. Ross is [a member of faculty](#) at the School of Public Health, Physiotherapy and Sports Science, University College Dublin, Ireland.

Greetings! I am undoubtedly one of the many researchers who have taken the editor, Will Hopkins, up on his statement in *New View of Statistics*: “Feedback wanted: if you can’t understand something here, it’s my fault. Email me.” Since our initial contact over email in February 2017, I have been working with Will in a mentoring relationship that has led to me developing high levels of proficiency in the use of the resources here at Sportsscience.

So, what is the purpose of having an associate editor? I have come on board primarily to extend the scope of the Sportsscience website. The readership of the journal site and users of its resources extend beyond athletic performance, so it is about time that the contributors to the site do too! Plus I suspect I am being groomed to take over the site.

I am particularly keen to develop a physical activity- and health-related strand of resources and meta-research here at Sportsscience. My own research is on physical education interventions

and assessment, and on sport and exercise science in childhood and adolescence more broadly. My experience here has led me to believe that there is work to be done in these fields to promote the more regular use of effect magnitudes and methods accounting for precision of estimation, such as MBI (or MBD, henceforth) and other approaches that prioritize interpreting the upper and lower confidence limits for an effect. Additionally, I suspect that the samples and challenges in research on physical activity and health are more diverse than those on competitive athletes, so I will be working to extend the resources in this new direction.

A first glance of an output associated with this new direction is already published here in the form of my [report](#) on the 23rd Annual Meeting of the European College of Sport Science (ECSS), which was hosted by my home institution in Dublin in July 2018. This year there will be three reports on the ECSS conference in Prague: Will’s usual on athletic performance, mine on physical activity and education in childhood and adolescence, and one from Lars Donath and his team on activity in adults, seniors and clinical populations. Watch this space.

Magnitude-Based Decisions

Will G Hopkins, Institute for Health and Sport, Victoria University, Melbourne, Australia. [Email](#). Reviewer: Ross D Neville, School of Public Health, Physiotherapy and Sports Science, University College Dublin, Dublin, Ireland. Sportsscience 23, i-iii, 2019 (sportsci.org/2019/inbrief.htm#decisions. Published June 2019. ©2019

Earlier this year I contacted the statistician Sander Greenland for clarification of a remark he had made in a [discussion about Bayesian priors](#) on the datamethods.org site. In the subsequent interactions, Greenland provided extensive advice on how to present MBI to a skeptical statistics community. He is opposed to the use of the term *inference*, unless it includes consideration not only of the sampling uncertainty in the

magnitude (regardless of the frequentist, Bayesian or other interpretation of the uncertainty) but also of all the other potential biases arising from violation of assumptions about sampling and the analytic model. He agrees that it would be appropriate to rebrand MBI as a method for making magnitude-based *decisions* (MBD). He also prefers *compatibility* to *confidence* limits and intervals, in the sense that the interval defines a range

of values interpreted as being compatible with the data and the statistical model (Greenland, 2019). *MBD* and *compatibility* have now been edited into the spreadsheets at *Sportsscience*.

The recent call by Greenland and his co-authors to retire statistical significance (Amrhein et al., 2019) raises an important question: if you can't use $p < 0.05$, how will you decide whether you have found something useful, important, or publishable in your sample? Some statisticians argue for leaving it to readers to make their own decisions, such that "most discovery claims would be replaced by description", as Greenland stated in [a recent tweet](#). Nevertheless, sport scientists often undertake research to decide if an intervention is implementable in their setting, and journal editors need to make decisions about adequate precision of effect magnitudes in manuscripts. For these scenarios I think the *MBD* method is a good answer. Alan Batterham and I have already provided cogent explanations of the decision process and of the ways erroneous decisions can arise. Furthermore we have done simulations to show that the error rates are acceptable (Hopkins and Batterham, 2016; Hopkins and Batterham, 2018).

Greenland also would like to see *MBD* couched in terms of "equivalence, minimal-effects, and non-inferiority hypothesis testing", to quote from one of his emails. The levels of the compatibility intervals in the clinical and non-clinical versions of *MBD*, and the disposition of the intervals relative to the thresholds for smallest important effects, should provide the hypothesis testing that will satisfy at least Greenland. The Bayesian probabilistic outcomes of *MBD* and the resulting decisions, whether derived with non-informative or informative priors (see below), would not change.

I was about to publish this item when a critique of *MBI* appeared in *Scandinavian Journal of Medicine and Science in Sports* (Sainani et al., 2019). Much of the critique reiterates what was already stated in *Medicine and Science in Sports and Exercise* last year (Sainani, 2018) and in a response (Sainani, 2019) to a rebuttal article (Hopkins and Batterham, 2018) and letter (Batterham and Hopkins, 2019). The two key points in the critique are that *MBI* is not Bayesian (or, if it is Bayesian, its implicit flat prior is "unrealistic"), and that the error rates with *MBI* are too high. The critique includes criticism of a spreadsheet downloaded from this site and used

by authors of an article published recently in the Scandinavian journal (Pamboris et al., 2019). I will now deal briefly with the two key points and with the criticism of the spreadsheet.

The [article](#) in this issue accompanying a new spreadsheet for Bayesian analysis (Hopkins, 2019) addresses the issue of whether *MBI* is Bayesian. In *MBI*, the usual frequentist confidence interval is interpreted in a Bayesian fashion as the likely range of the true value; equivalently, the associated *t* distribution is interpreted as a probability distribution of the true value. Many previous authors have promoted this interpretation (e.g., most recently, Albers et al., 2018), but it requires an assumption of a "flat" prior—no prior belief or information about the true effect—an assumption that some statisticians regard as implying that the true value could have unrealistic huge values. In the examples shown in the Bayesian spreadsheet, you will see that a realistic *weakly* informative prior makes no practical difference to compatibility (confidence) intervals and magnitude-based decisions with the sometimes unavoidably small sample sizes that sport scientists have to use, and, of course, with any larger sample sizes. It follows that a flat prior is effectively a realistic prior for such studies, and therefore that the probabilistic statements of *MBI* are legitimate Bayesian.

The claim for high error rates with *MBI* now rests on evidence that *MBI* is widely misused by researchers, who apparently treat *possibly* and *likely substantial* as *definitively substantial*. The error rates with such misuse are indeed unacceptable, but this kind of misuse can easily be corrected in the review process. The authors of the present and previous critique of *MBI* also failed to point out that, for clinically relevant effects, researchers should and do consider *possibly* or *likely beneficial* as potentially *implementable* (after a cost-benefit analysis), and that even here the resulting Type-1 error rates are acceptable. I won't address again the claims of high error rates based on null-hypothesis significance testing, since I side with those calling for the retirement of statistical significance. The null hypothesis has no place in the real world.

Pamboris et al. (2019) performed a crossover to quantify the acute effects of two kinds of stretching on exercise-induced changes in neuromuscular variables. The authors used the spreadsheet for a parallel-groups trial, when instead they should have used the spreadsheet for a pre-

post crossover. This mistake should have been picked up in the review process and indeed by the authors of the critique. Instead, they suggested that the spreadsheet was inadequate for not "correctly handling correlated observations"; even more unjustifiably, "a basic statistical error described in any introductory statistics course" was then ascribed to "the MBI approach". The theory underlying the controlled-trial and other spreadsheets is well documented in accompanying articles going back 16 years, and the spreadsheets give the same answers as mixed modeling with the Statistical Analysis System, including estimates and compatibility limits for individual responses. In most cases the SAS code is provided. These are all basic analyses that I did not consider worth publishing in statistical journals, especially when I had validated them with SAS programs.

The authors of the critique also called into question the use of log transformation in the spreadsheet, which "can make results difficult to interpret" and "can also make it harder to check simple numbers". But any effects (and errors) that are more likely to be uniform when expressed in factor or percent units should be analyzed with logs, including effects with the majority of dependent variables in exercise and sport science: all those involving time, distance, force, power, work and concentration, where only positive values are possible. Confidence limits in \pm form for percent effects also came in for criticism, but I indicated clearly in all the spreadsheets that this form is an approximation (which I espouse to reduce digital clutter), and the inferential statistics are obviously not derived with these limits. Concerns about inconsistencies with the other inferential statistics are therefore misplaced.

The authors of the critique made a good point about "black box" approaches, which allow users to analyze data and make mistakes through lack of understanding of statistical principles. Whether the spreadsheet deserves to be categorized as a black box is debatable, considering all formulae are visible in the cells of the spreadsheet, and many cells contain extensive explanatory comments. In alerting me to their critique,

one of my colleagues made the following opposite remark: "I have seen students, supervisors and well established researchers put data into SPSS, GraphPad, SigmaPlot, etc., and take the result for granted, as with your spreadsheets. So a call for end-user education rather than criticism of your methods seems more reasonable."

- Albers CJ, Kiers HA, van Ravenzwaaij D (2018). Credible confidence: a pragmatic view on the frequentist vs Bayesian debate. *Collabra: Psychology* 4, <https://collabra.org/articles/10.1525/collabra.149>
- Amrhein V, Greenland S, McShane B (2019). Retire statistical significance. *Nature* 567, 305-307
- Batterham AM, Hopkins WG (2019). The problems with "The Problem with 'Magnitude-based Inference'". *Medicine and Science in Sports and Exercise* 51, 599
- Greenland S (2019). Valid P-values behave exactly as they should: Some misleading criticisms of P-values and their resolution with S-values. *The American Statistician* 73, 106-114
- Hopkins WG, Batterham AM (2016). Error rates, decisive outcomes and publication bias with several inferential methods. *Sports Medicine* 46, 1563-1573
- Hopkins WG, Batterham AM (2018). The vindication of magnitude-based inference. *Sportsscience* 22, 19-29
- Hopkins WG (2019). A spreadsheet for Bayesian posterior compatibility intervals and magnitude-based decisions. *Sportsscience* 23, 5-7
- Pamboris GM, Noorkoiv M, Baltzopoulos V, Mohagheghi AA (2019). Dynamic stretching is not detrimental to neuromechanical and sensorimotor performance of ankle plantarflexors. *Scandinavian Journal of Medicine and Science in Sports* 29, 200-212, <https://onlinelibrary.wiley.com/doi/pdf/10.1111/sms.13321>
- Sainani KL (2018). The problem with "magnitude-based inference". *Medicine and Science in Sports and Exercise* 50, 2166-2176
- Sainani KL (2019). Response. *Medicine and Science in Sports and Exercise* 51, 600
- Sainani KL, Lohse KR, Jones PR, Vickers A (2019). Magnitude-Based Inference is not Bayesian and is not a valid method of inference. *Scandinavian Journal of Medicine and Science in Sports* (in press), doi: 10.1111/sms.13491