Title: Estimation versus falsification approaches in sport and exercise science.

Submission Type: Short communication

Authors: 1. Michael Wilkinson
2. Edward M Winter

Affiliation: 1. Department of Sport, Exercise and Rehabilitation
Faculty of Health and Life Sciences
Northumbria University
Northumberland Building
Newcastle-upon-Tyne
NE1 8ST
ENGLAND
Email: mic.wilkinson@northumbria.ac.uk
2. Centre for Sport and Exercise Science
Academy of Sport and Physical Activity
Sheffield Hallam University
Collegiate Crescent
Sheffield
S10 2BP
Email: e.m.winter@shu.ac.uk

Abstract word count: 164

Text only word count: 945

Number of figures = 0; number of tables = 0

Abstract

There has been a recent resurgence in debate about methods for statistical inference in science. The debate addresses statistical concepts and their impact on the value and meaning of analyses' outcomes. In contrast, philosophical underpinnings of approaches and the extent to which analytical tools match philosophical goals of the scientific method have received less attention. This short piece considers application of the scientific method to 'what-is-the-influence-of x-on-y' type questions characteristic of sport and exercise science. We consider applications and interpretations of estimation versus falsification based statistical approaches and their value in addressing how much x influences y, and in measurement error and method agreement settings. We compare estimation using magnitude based inference (MBI) with falsification using null hypothesis significance testing (NHST), and highlight the limited value both of falsification and NHST to address problems in sport and exercise science. We recommend adopting an estimation approach, expressing the uncertainty of effects of x on y, and their practical/clinical value against pre-determined effect magnitudes using MBI.

Keywords

Philosophy of science; statistical inference; falsification; estimation.

Since the times of William Sealey Gosset (1876-1937) and Ronald Aylmer Fisher (1890-1962), there has been debate about approaches to statistical inference. In the last three decades, debate has undergone a resurgence with emphasis on shortcomings of conventional null hypothesis significance testing (NHST) (Cohen, 1994; Wasserstein and Lazar, 2016) and implications for sport, exercise and medicine research (Hopkins *et al*., 2009; Wilkinson, 2014). Arguments have largely been about statistical concepts and how they impact on the meaning and value of outcomes of analyses. An aspect of the debate that has received less attention is the philosophical underpinning of approaches discussed and how well these tools match the application of the scientific method to the types of problem characteristic of sport and exercise science. In this short piece, we consider the scientific method, the types of research question commonly addressed, and applications and interpretations of outcomes of statistical analyses that might best match the questions we ask.

The scientific method, falsification and NHST

Science is a systematic process of addressing questions, characterised by inductive reasoning from observations in experiments, to general theories that attempt to explain observations and predict future events. According to philosopher Karl Popper (1934, 1969), the value of such theories lies in their ability to generate specific predictions (hypotheses) that can be compared against experimental results and, potentially, be shown to be wrong. Popper suggested that specific predictions at odds with experimental data falsify the prediction in a pass-fail manner, and should lead to amendment or abandonment of the parent theory. It is through such amendments or abandonment of theories by falsification that, Popper claimed, science and knowledge advanced. He further wrote that theories unable to generate potentially falsifiable and specific predictions, were not scientific (Popper, 1934).

The attempted falsification of hypotheses by experiment has been suggested to be an 'ideal' application of the scientific method (Ladyman, 2002). It also seems to support the statistical approach of NHST whereby researchers strive to falsify their null hypothesis in a pass-fail manner.

Criticisms of the falsification approach

As a theory, Popper's thesis must itself be subject to falsification. In his classic text 'The Structure of Scientific Revolutions' (1962), Thomas Kuhn systematically dismantled Popper's thesis using a detailed historical account of scientific progress in disciplines including physics and astronomy. The 'paradigm shifts' Kuhn described were completely at odds with the incremental progress suggested by Popper. Moreover, Kuhn found no evidence that theories are simply abandoned, as Popper suggested, in the face of a falsifying observation or an accumulation thereof. The dichotomous decision accompanying falsification is mirrored by, and has been widely criticised in, NHST (Cohen, 1994; Batterham and Hopkins, 2006; Ziliak

and McCloskey, 2008). Perhaps it is time to consider a philosophical approach that improves standard scientific practice, and a statistical approach that does likewise.

Research questions in sport and exercise science

Most, if not all research questions in sport and exercise science take the form 'does x predict/correlate with y, or what is the effect of x on y? In fact, both can be expressed simply as 'does x influence y'? This can be addressed with a yes/no approach (NHST) or with a 'how much?' approach (interval estimation). The former requires a specific threshold and definition of what 'yes' and 'no' mean. This should be an effect magnitude of practical/clinical value if NHST is performed correctly. Such effect magnitudes are rarely stated in studies using NHST. Instead, an arbitrarily chosen threshold of $P < 0.05$, is considered the pass-fail criterion, where P is the probability of an effect magnitude as big or bigger than that observed occurring in the long run, if the true population effect size were zero. This process leads to decisions with high error rates (Colquhoun, 2014; Hopkins and Batterham, 2016) and without practical or clinical context (Batterham and Hopkins, 2006). Even if a threshold with real-world context is specified, the uncertainty of observed/sample effect sizes (because say, of sampling error) cloud a simple dichotomous decision about the 'true' population effect. The value of any philosophical and statistical approach with an absolute dichotomous outcome is therefore questionable.

These problems are negated by an estimation/how much? approach to science and statistical analysis. Specific and absolute predictions/thresholds of effect or are not required here. All that is needed is a sense of effect magnitude that is of practical/clinical importance. The researcher simply sets out to estimate the magnitude of the influence of x on y, expressing the uncertainty around the estimate with a confidence interval, and interpreting the interval in the context of a practically/clinically worthwhile effect. This approach is simple, informative, less prone to error (Hopkins and Batterham, 2016) and improves research activity in sport and exercise science. It is encapsulated by MBI.

Measurement error and/or method agreement studies present another problem for the falsification approach and accompanying NHST. Such studies are crucial to rigorous measurement and interpretation of interventions, but the goal is simply to estimate variability between repeated measurements under standard conditions with the same tool, or between different measurement tools. This variability is not known before a study. A predicted falsifiable outcome and pass-fail approach is therefore unhelpful in these settings. An estimation approach however, is well suited to the philosophical and analysis goals.

In summary, Popper's falsification thesis and the NHST approach (believed by many to embrace falsification), are of limited value for tackling questions characteristic of sport and exercise science. Some of these issues have been addressed in editorials in this Journal (Winter, 2008; Winter *et al*., 2014) and in detail elsewhere (Batterham and Hopkins, 2006).

The philosophical goals of sport and exercise science are better served by an estimation approach and statistical methods that embody this, namely MBI.

References

Batterham, A.M. & Hopkins, W.G. (2006). Making meaningful inferences about magnitudes. *International Journal of Sports Physiology and Performance,* 1, 50-57.

Cohen J. (1994). The earth is round (p< .05). *American Psychologist*, 49, 997–1003.

Colquhoun, D. (2014). An investigation of the false discovery rate and the misinterpretation of *p*-values. *Royal Society Open Science,* 1, 140216.

Hopkins, W.G., Marshall, S.W., Batterham, A.M. and Hanin, J. (2009). Progressive statistics for studies in sports medicine and exercise science. *Medicine and Science in Sports and Exercise*, 41, 3-13.

Hopkins, W.G. & Batterham, A.M. (2016). Error Rates, Decisive Outcomes and Publication Bias with Several Inferential Methods. *Sports Medicine,* 46(10), 1563-1573.

Kuhn, T. S. (1962). *The Structure of Scientific Revolutions.* Chicago: University of Chicago Press.

Ladyman, J. (2002). *Understanding Philosophy of Science.* London: Routledge.

Popper, K. (1934) *The Logic of Scientific Discovery.* London: Hutchinson.

Popper, K. (1969) *Conjectures and Refutations*. London: Routledge and Kegan Paul.

Wasserstein, R.L. and Lazar, N.A. (2016). The ASA's statement on p-values: context, process, and purpose. *The American Statistician*, 70(2), 129-133.

Wilkinson, M. (2014). Distinguishing between statistical significance and practical / clinical meaningfulness using statistical inference. *Sports Medicine*, 44, 295-301.

Winter, E.M. (2008). Use and misuse of the term "significant". *Journal of Sports Sciences*, 26, 429-430.

Winter, E.M., Abt, G.A. and Nevill, A.M. (2014). Metrics of meaningfulness as opposed to sleights of significance. *Journal of Sports Sciences*, 32, 901-902.

Ziliak, S.T. & McCloskey, D.N. (2008). *The Cult of Statistical Significance*. Ann Arbor: University of Michigan Press.