Title: Testing the null hypothesis: the forgotten legacy of Karl Popper?

Submission Type: Letter to the Editor

Abstract word count: 188

Text only word count: 999

Number of figures = 0; number of tables = 0

Abstract

Testing of the null hypothesis is a fundamental aspect of the scientific method and has its basis in the falsification theory of Karl Popper. Null hypothesis testing makes use of deductive reasoning to ensure that the truth of conclusions are irrefutable. In contrast, attempting to demonstrate the new facts on the basis of testing the experimental or research hypothesis makes use of inductive reasoning and is prone to the problem of the Uniformity of Nature assumption described by David Hume in the eighteenth century. Despite this issue and the well documented solution provided by Popper's falsification theory, the majority of publications are still written such that they suggest the research hypothesis is being tested. This is contrary to accepted scientific convention and possibly highlights a poor understanding of the application of conventional significance-based data analysis approaches. Our work should remain driven by conjecture and attempted falsification such that it is always the null hypothesis that is tested. The write up of our studies should make it clear that we are indeed testing the null hypothesis and conforming to the established and accepted philosophical conventions of the scientific method.

As a PhD student, my supervisor encouraged me to acquire a full understanding of the philosophy of science and to understand the basis and logic underpinning the scientific method. At the time, I admit I could not see the value of this endeavour or appreciate how it would impact on my research. As a dutiful student, I did as suggested and was quickly hooked on the intricacies and historical development of the epistemological approach we call science. Gaining an appreciation of the philosophy underpinning the scientific method and more importantly, understanding its limitations, made me a far more assiduous researcher. Recent reviewer comments on one of my own manuscripts, and a perusal of the latest issue of the Journal of Sports Sciences has brought to light that others might perhaps not be as well versed in the underpinning philosophy of our method, or have forgotten about it.

In response to a line of text in the introduction section of my recently reviewed manuscript reading "in light of previous literature, we tested the following null hypotheses...", the reviewer requested an alteration to state that the experimental / alternative hypotheses were to be tested. My reading in the philosophy of science and specifically of falsification theory, so eloquently discussed in the classical text *The Logic of Scientific Discovery* (Popper, 1980), suggested that I was correct and the reviewer was perhaps wrong. Establishing facts and theories by attempting to demonstrate the 'truth' of the alternative hypothesis makes use of inductive reasoning and was characteristic of early scientific endeavour as outlined in Sir Francis Bacon's book (1620) *Novum Organum* (translated as new tool) (Ladyman, 2008). However, in his book *An Enquiry Concerning Human Understanding,* originally published in 1748, Scottish philosopher David Hume correctly pointed out that conclusions based on inductive reasoning were flawed as no number of confirmatory observations of a theory or prediction could 'prove' a theory or prediction correct (Hume, 1963). To do so required an assumption that the same results would occur again on all future tests. Hume called this problem the Uniformity of Nature assumption. It casts doubt on any and all scientific conclusions based on inductive reasoning and it has yet to be resolved. Popper's solution and lasting contribution to science was to suggest that while inductive reasoning was necessary to form theories from observations, deductive reasoning could be used to derive predictions from theories and to falsify or refute such predictions. If a prediction of a theory withstood falsification, it remained useful. The strength of deductive reasoning is that the truth of a conclusion derived in this way is irrefutable if the preceding facts are known to be correct. An often used and simple example can simultaneously highlight the problem of inductive reasoning and the strength of deductive reasoning. Suppose that a scientist hypothesises that all swans are white. After careful observation to test this hypothesis and no sightings of a non-white swan, the scientist concludes that all swans are indeed white. This is inductive reasoning. Even if the observations were accurate, the conclusion might still be false as the scientist assumes that all other swans that had yet to be observed would also be white. Applying Popper's falsification theory, the scientist could formulate a null hypothesis that all swans are not white. Suppose that he subsequently spots a non-white swan at place *x* and time *t*, it logically follows by deduction that all swans are not white and the truth of that conclusion is irrefutable. Hence knowledge and accepted facts could be advanced by conjecture and attempted refutation / falsification (Popper, 1969). An excellent and thorough discussion of the problems of induction and Popper's elegant solution as applied to the sport and exercise sciences is provided by McNamee (2005).

In the accepted scientific method, the falsity of the experimental hypothesis (H_1) is expressed in the form of the null hypothesis (H_0) and it is the latter that is subjected to scrutiny using probability-based statistics. Assessing the probability that a particular effect observed in a sample could occur if the null hypothesis was in fact the true population effect is the basis of significance tests that remain the conventional approach of data analysis in most sport and exercise science publications. However, problems with this statistical approach have been highlighted and other metrics including effect size, confidence intervals and magnitude-based inference have been proposed as better alternatives (Hopkins *et al.*, 2009). While the benefits of these developments are clear, proponents call for the abandonment of hypothesis testing altogether. Here we must proceed with caution, for as Popper and Hume before him rightly pointed out, any claim to the existence of an effect cannot be made unless it is first shown that a situation of no effect is untenable. It is suggested that null hypothesis tests be carried out to first establish that a population effect is in fact unlikely to be zero. This step can then be followed by a confidence-interval based approach that estimates what the magnitude of effect might plausibly be. Finally, a probability associated with the likelihood of the population effect exceeding an *a priori* smallest meaningful effect can be calculated. This sequence ensures that the logical reasoning underpinning a conclusion of the existence of some effect remains robust and we do not suffer the pitfalls inherent in inductive reasoning. It must be noted however that, even when we follow this sequence, our confidence in rejecting H_0 and in our estimates of true but unknown population effects are still made with a degree of uncertainty that depends on aspects of the study such as sample selection, sample size, consistency of response and accuracy and sensitivity of measures (and therefore statistical power). In summary, our work should remain driven by conjecture and attempted falsification such that it is always the null hypothesis that is tested. The write up of our studies should make it clear that we are indeed testing the null hypothesis and conforming to the established and accepted philosophical conventions of the scientific method.

References

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**(1), 3-12.

Hume, D. (1963) *An Enquiry Concerning Human Understanding.* Oxford: Oxford University Press.

Ladyman, J. (2008) *Understanding Philosophy of Science.* Oxon: Routledge.

McNamee, M. (2005). Positivism, Popper and Paradigms: and introductory essay in the philosophy of science. In M. McNamee (Ed.), *Philosophy and the Sciences of Exercise, Health and Sport: Critical perspectives on research methods* (pp. 1-21). Oxon: Routledge.

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

Popper, K. (1980) *The Logic of Scientific Discovery* (10th Ed). London: Hutchinson.