Validating new discoveries in sports medicine: we need FAIR play beyond p values

Chris Bleakley 1,2, James M Smoliga2

There is concern that a large proportion of scientific research is based on false-positive, non-replicable conclusions. As most experimental research in sports medicine is based on frequentist reasoning, p values have been at the centre of knowledge claims and new discoveries within this field. But many researchers and clinicians are unable to define or accurately interpret p values. Common misconceptions are that p values represent ‘the probability that the null hypothesis is true’ or ‘the probability that the hypothesis being tested is true’. In effect, p values only quantify the chances of getting the observed data (on the assumption that the null hypothesis is true) and therefore cannot exclusively inform clinical decision making. This editorial presents FAIR: a four-item approach to help validate new discovery in sports medicine.

FALSE-POSITIVE RISK

False-positive risk (FPR) is ‘the probability of observing a statistically significant p-value and declaring that an effect is real, when it is not’. Crucially, a study’s FPR can be high, even when the corresponding p values are low. In a systematic audit of high-quality randomised controlled trials (RCTs) in sports physiotherapy, 18% of ‘statistically significant’ findings had a 50% chance of false discovery (claiming a treatment effect is real when it is not). FPR calculation is underpinned by Bayes’ theorem, whereby information from two sources (the prior probability of treatment success and the data from the experiment) are combined to provide a ‘posterior probability’ of treatment success. When appraising experimental research, we can reverse this logic using the data to estimate the prior probability of treatment success, while being cognisant that a neutral prior (a 50:50 chance of treatment success) is perhaps the largest that can be legitimately assumed. For example, an experimental study (n=20 per group) reporting a large effect size (1.1) and a p value of 0.049 corresponds to a prior probability of 97%—if we assume an FPR of 5%. We should be cautious of such an inflated prior as it suggests either: A). the experiment was potentially unnecessary as researchers were almost certain of treatment success at the study’s inception; or B). that the FPR exceeds the set threshold (eg, 5%) and there is elevated risk of false discovery.

A PRIORI REGISTRATION

Currently, only one in three RCTs in sports physiotherapy is prospectively registered. A priori registration of clinical trials ensures that key study details, including primary outcomes, are made public prior to analysis. Unregistered trials carry a higher risk of false discovery, due to unplanned multiple testing, selected reporting and confirmation bias. Registration can help to control the ‘degrees of freedom’ a researcher has when making small but important decisions regarding data analysis and reporting. The corollary is that positive conclusions from prospectively registered RCTs should hold most weight, with positive findings from unregistered studies considered as exploratory or even hypothesis generating.

CLINICAL IMPORTANCE

P values do little to indicate the clinical importance of observed treatment effects. Effect measures are more intuitive, but standard scores (eg, standardised mean difference) do not provide immediate clinical context. Therefore, legitimate clinical importance can only be determined by framing the difference in means (±CIs) with relevant minimal detectable change (MDC) and minimal clinical important difference (MCID) thresholds. MDC represents ‘the amount of change (in the outcome) that must be observed before it is considered above the bounds of measurement error’; and MCID represents ‘the smallest change (in the outcome) that would be important to patients’. These thresholds are commonly overlooked, and a 2018 audit found that just 7% of orthopaedic researchers

1School of Health Sciences, University of Ulster at Jordanstown, Newtownabbey, UK
2Congdon School of Health Sciences, High Point University, High Point, North Carolina, USA
Correspondence to Dr Chris Bleakley, University of Ulster at Jordanstown, Newtownabbey, Antrim, UK; chrisbleakley77@gmail.com

Figure 1 Statistical significance vs clinical importance. Both A and B reported statistically significant changes (p<0.05).6,7 Dots (and whiskers) represent mean change scores (95% CIs). Ankle dorsiflexion measured through a weight bearing lunge test (cm); MDC=1.9 cm; MCID=2 cm. MCID, minimal clinical important difference; MDC, minimal detectable change.
Figure 1 shows data from two experimental studies, each reporting statistically significant changes in ankle dorsiflexion post intervention (p<0.05). Despite this, the treatment effects observed in study A cannot be differentiated from measurement error. Although study B reports a larger average effect, most of the observed changes do not reach the threshold for clinical importance (MDC+MCID) and are unlikely to be meaningful to patients.

REPLICATION
The replication crisis is a ubiquitous and complex problem across all of science. Sports medicine has been slower to react compared with other fields of medicine; currently, the volume of research in this field which has been successfully corroborated through replication is unclear. FAIR reminds clinicians and researchers that independent replication underpins scientific discovery and that it is presumptuous to conclude treatment effectiveness based on a single significant result.

SUMMARY
Time restraints and lack of training are cited as common barriers preventing clinicians from fully engaging in the evidence base. P value thresholds (is p<0.05?) offer a fast but ultimately limited method for determining clinical effectiveness. Although there are many other aspects of trial design and reporting that can increase the risk of false discovery, FAIR is presented as a preliminary concept to help clinicians disentangle true-positive from potentially false-positive claims within sports medicine.

Contributors
Both authors certify that they have participated sufficiently in the work to take public responsibility for the content, including participation in the concept, design, analysis, writing or revision of the manuscript. CB and JMS were involved in the concept, design and writing. Both authors were involved in final submission and revision of the manuscript.

Funding
The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests
None declared.

Patient consent for publication
Not required.

Provenance and peer review
Not commissioned; externally peer reviewed.

To cite
Bleakley C, Smoliga JM. Br J Sports Med Epub ahead of print: [please include Day Month Year]. doi:10.1136/bjsports-2019-101797

Accepted 8 June 2020
Br J Sports Med 2020;0:1–2. doi:10.1136/bjsports-2019-101797

ORCID
Chris Bleakley http://orcid.org/0000-0001-9032-9691

REFERENCES
1. Ioannidis JPA. Why most published research findings are false. PLoS Med 2005;2:e124.
2. Colquhoun D. The False Positive Risk: A Proposal Concerning What to Do About p -Values. Am Stat 2019;73:192–201.
3. Bleakley CM, Reijgers J, Smoliga JM. Many high quality RCTs in sports physical therapy are making false positive claims of treatment effect: a systematic survey. J Orthop Sport Physio. In Press.
4. Forstmeier W, Wagenmakers E-J, Parker TH. Detecting and avoiding likely false-positive findings - a practical guide. Biol Rev Camb Philos Soc 2017;92:1941–68.
5. Copay AG, Eyberg B, Chung AS, et al. Minimum clinically important difference: current trends in the orthopaedic literature, part II: lower extremity: a systematic review. JBJS Rev 2018;6:e2.
6. Driller M, Mackay K, Mills B, et al. Tissue flossing on ankle range of motion, jump and sprint performance: a follow-up study. Phys Ther Sport 2017;28:29–33.
7. Marrón-Gómez D, Rodríguez-Fernández Ángel L, Martín-Urrutia JA. The effect of two mobilization techniques on dorsiflexion in people with chronic ankle instability. Phys Ther Sport 2015;16:10–15.