



ELSEVIER

Contents lists available at ScienceDirect

Journal of Science and Medicine in Sport

journal homepage: www.elsevier.com/locate/jsams

Journal of Science and Medicine in Sport

Original research

The bias for statistical significance in sport and exercise medicine

David N. Borg^{a,*}, Adrian G. Barnett^a, Aaron R. Caldwell^b, Nicole M. White^a, Ian B. Stewart^c^a Australian Centre for Health Services Innovation and Centre for Healthcare Transformation, School of Public Health and Social Work, Queensland University of Technology, Australia^b United States of America^c School of Exercise and Nutrition Sciences, Queensland University of Technology, Australia

ARTICLE INFO

Article history:

Received 11 October 2022

Received in revised form 1 March 2023

Accepted 2 March 2023

Available online xxxx

Keywords:

Confidence interval

Estimation

Misconduct

p-hacking

Publication bias

Selective reporting

ABSTRACT

Objectives: We aimed to examine the bias for statistical significance using published confidence intervals in sport and exercise medicine research.

Design: Observational study.

Methods: The abstracts of 48,390 articles, published in 18 sports and exercise medicine journals between 2002 and 2022, were searched using a validated text-mining algorithm that identified and extracted ratio confidence intervals (odds, hazard, and risk ratios). The algorithm identified 1744 abstracts that included ratio confidence intervals, from which 4484 intervals were extracted. After excluding ineligible intervals, the analysis used 3819 intervals, reported as 95 % confidence intervals, from 1599 articles. The cumulative distributions of lower and upper confidence limits were plotted to identify any abnormal patterns, particularly around a ratio of 1 (the null hypothesis). The distributions were compared to those from unbiased reference data, which was not subjected to *p*-hacking or publication bias. A bias for statistical significance was further investigated using a histogram plot of *z*-values calculated from the extracted 95 % confidence intervals.

Results: There was a marked change in the cumulative distribution of lower and upper bound intervals just over and just under a ratio of 1. The bias for statistical significance was also clear in a stark under-representation of *z*-values between -1.96 and $+1.96$, corresponding to *p*-values above 0.05.

Conclusions: There was an excess of published research with statistically significant results just below the standard significance threshold of 0.05, which is indicative of publication bias. Transparent research practices, including the use of registered reports, are needed to reduce the bias in published research.

© 2023 The Author(s). Published by Elsevier Ltd on behalf of Sports Medicine Australia. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Practical implications

- An examination of 3819 confidence intervals showed that there was a large excess of published research with statistically significant results, just below the standard significance threshold of 0.05.
- Transparent research practices, including the use of registered reports, are needed to reduce the bias in published research.
- There is an urgent need for peer reviewers, editors, and journals to direct the rewards of publication away from statistical significance and onto scientific rigor.

1. Introduction

Every sport and exercise medicine researcher should be aware that a statistically significant result is more likely to get published.^{1–3} The selective publishing of statistically significant results has encouraged poor

practices, including *p*-hacking, the generation of post hoc hypotheses, and data fabrication.^{4–6} The bias toward the publication of significant results has also distorted evidence for scientific claims, with many non-statistically significant findings never making it to publication.⁷ In exercise medicine, the focus on statistical significance has been shown to bias a researcher's perceptions and decision making during a study.¹ For example, the decision to collect more data when a result does not reach the specified significance threshold—usually a *p*-value of 0.05. In defense of researchers, significance-seeking behaviors may not always be overt, and can occur despite seemingly reasonable decisions being made.^{8,9}

Bias around statistical significance is often examined using *p*-curves.⁵ A *p*-curve is a plot of the distribution of reported *p*-values that fall below a chosen threshold for defining statistical significance, most commonly 0.05. An excess of *p*-values that fall just below the chosen threshold is statistically implausible and signals publication bias.

There have been recent calls for researchers to replace *p*-values with confidence intervals in order to reduce the bias promoted by the overuse of *p*-values.^{10–12} However, there is no empirical evidence that emphasizing confidence intervals over *p*-values reduces *p*-hacking and publication bias.^{13–15}

* Corresponding author.

E-mail address: dn.borg@qut.edu.au (D.N. Borg).

No previous study has used confidence intervals to examine bias regarding statistical significance in the sport and exercise medicine literature. We aimed to assess the presence of bias around the statistical significance threshold using ratio confidence intervals (odds ratios, hazard ratios, risk ratios) as these can be accurately extracted from published papers by automated tools. We hypothesized that there would be a marked change in the cumulative distribution of upper and lower bound intervals near a ratio of 1, which is the null hypothesis of no difference on a ratio scale.

2. Methods

We used a validated text-mining algorithm^{13,16} to extract confidence intervals (see “Ratio confidence intervals” box) from the abstracts of articles published in 18 sports and exercise medicine journals between 2002 and 2022, that are indexed in *MEDLINE* (Supplement 1). The text-mining algorithm recognizes the typical ways in which authors report statistical ratios (odds ratios, hazard ratios, risk ratios), and identifies and extracts the confidence intervals reported with a ratio. No ethical approval for the study was needed as we used publicly available data that is published to be read and scrutinized.

Eighteen journals were selected from a list of the top 100 journals in the subject area of Physical Therapy, Sports Therapy and Rehabilitation on Scimago. We chose any journal that included the word ‘medicine’ in the name and appeared in *MEDLINE*. The extraction was restricted to original articles and reviews. Our focus was on journals that appeared in *MEDLINE* over the past two decades, but to increase the sample size we also included three journals that appeared in *MEDLINE* after 2002 and continued to 2022. These three journals were: *Research in Sports Medicine* (appears from 2005 onwards), *Sports Medicine and Arthroscopy Review* (2006 onwards) and *European Journal of Physical and Rehabilitation Medicine* (2008 onwards).

The text-mining algorithm was designed to recognize regular expressions that authors use to report statistical ratios. For example, “OR = 0.42, 95 % CI = 0.16–1.13”, where ‘OR’ is the odds ratio and ‘CI’ the confidence interval. The text-mining algorithm has previously been used to extract ratio confidence intervals to identify reporting errors¹⁶ and to investigate bias in ratio confidence intervals in the medical literature.¹³ In the current study, the text-mining algorithm was highly accurate with a true positive percentage of 99 % in a random sample of 100 abstracts. In the one missed observation, it was unclear whether the reported interval was an interquartile range or a confidence interval.

Confidence intervals were excluded from the analysis when: there was a boundary violation, that is, when the ratio point estimator was outside the confidence interval; the lower bound was below zero, which is not possible for ratios; and when the level of confidence interval was not reported.

Graphical summaries were used to examine the presence of bias in the distribution of intervals, particularly around the significance threshold, that is, a ratio of 1 (see “Ratio confidence intervals” box). The cumulative distributions of lower and upper bounds for all confidence intervals were plotted to highlight changes without the need for smoothing.

For comparison, we plotted the cumulative distributions alongside those generated from an unbiased reference dataset, which had not been subjected to *p*-hacking or publication bias.¹⁸ The unbiased dataset contains results from thousands of analyses performed on a large database of insurance claims, where all possible pairs of 17 depression treatments were compared using 22 outcomes. The results were given as hazard ratios and 95 % confidence intervals, with nearly 18,000 ratio confidence interval pairs produced. The results did not show an absence of small or non-statistically significant effects, and there was no sharp cut-off at a *p*-value of 0.05. These results provide a reference for the shape of the distribution of ratio confidence intervals if all study results were published and no bias was present.¹³ To compare our results to the field of medicine, we also plotted the cumulative distributions of the

extracted intervals against the results (abstracts only) published by Barnett & Wren.¹³

We plotted the cumulative distributions in 5-year blocks to investigate whether there was any change in the lower and upper intervals over time. We used 5-year blocks because the sample size was insufficient for yearly cumulative distributions.

A bias for statistical significance was further investigated using a histogram plot of *z*-values calculated from the extracted 95 % confidence intervals.¹⁹ For each confidence interval a *z*-value was calculated using the equation: $z = \log(\mu) / se$, where ‘ μ ’ is the mean estimate and ‘*se*’ is the standard error.

All analyses were undertaken in R (version 4.1.3),²⁰ using the packages *tidyverse*,²¹ *rentrez*,²² *XML*²³ and *naniar*.²⁴ The data and R code used to produce our results are available at <https://doi.org/10.5281/zenodo.7102324>. The code was adapted from <https://github.com/agbarnett/intervals>.

3. Results

Abstracts from 48,390 unique articles, published in 18 sports and exercise medicine journals between 2002 and 2022, were searched for ratio confidence interval pairs. The text-mining algorithm identified 1744 unique abstracts from 16 of the 18 journals that included ratio confidence intervals, from which 4484 intervals were extracted. A list of the journals searched, and the number of intervals extracted from these journals is in Supplement 1.

We removed interval pairs due to a boundary violation ($n = 104$; 2.3 %), a negative lower bound ($n = 14$; 0.3%), or a missing level of confidence ($n = 508$; 11.3 %), leaving 3858 interval pairs. In terms of missing data, the percentage of intervals missing the level of confidence decreased over time (Supplement 2 Panel A) and was as high as 26.3 % for one journal (Supplement 2 Panel B). Five journals had over 20 % intervals missing the level of confidence interval. When the level of confidence was provided, most intervals were given as 95 % confidence intervals ($n = 3819/3858$; 99%), with 90 % ($n = 30/3858$; 0.8 %) and 99 % ($n = 9/3858$; 0.2 %) intervals also reported.

Focusing on 95 % confidence intervals, 3819 interval pairs were extracted from 1599 articles. The cumulative distribution of these 3819 intervals showed that there was an excess of statistically significant results, with a clear inflection point in the distribution of lower bounds just over a ratio of 1, and to a lesser extent, upper bounds just below 1 (Fig. 1). This distinct distributional pattern was very similar to that observed in medical research (Fig. 1). The excess of statistically significant results has changed little over time (Supplement 3).

The excess of statistically significant results was clearly highlighted by the marked under-representation of *z*-values between -1.96 and $+1.96$, corresponding to *p*-values greater than 0.05, which is the commonly used significance threshold (Fig. 2). Fig. 2 clearly shows the enormous absence of published non-statistically significant results. We do not know the exact shape that the distribution would have in the absence of reporting and publication bias, but it would be far smoother than the observed distribution, and the mean and mode would be close to zero. An unbiased distribution would not necessarily be symmetric, as researchers may prefer to report differences as positive numbers.

4. Discussion

We used a validated text-mining algorithm to extract over 4000 ratio confidence intervals from nearly 1700 sport and exercise medicine articles between 2002 and 2022. We plotted the cumulative distribution of lower and upper 95 % confidence interval bounds to identify whether there were any abnormal changes in the distributions around the null hypothesis ratio of 1, which could be indicative of bias. As expected, there was a large excess of published research with statistically significant results, just below the standard significance threshold of 0.05. This

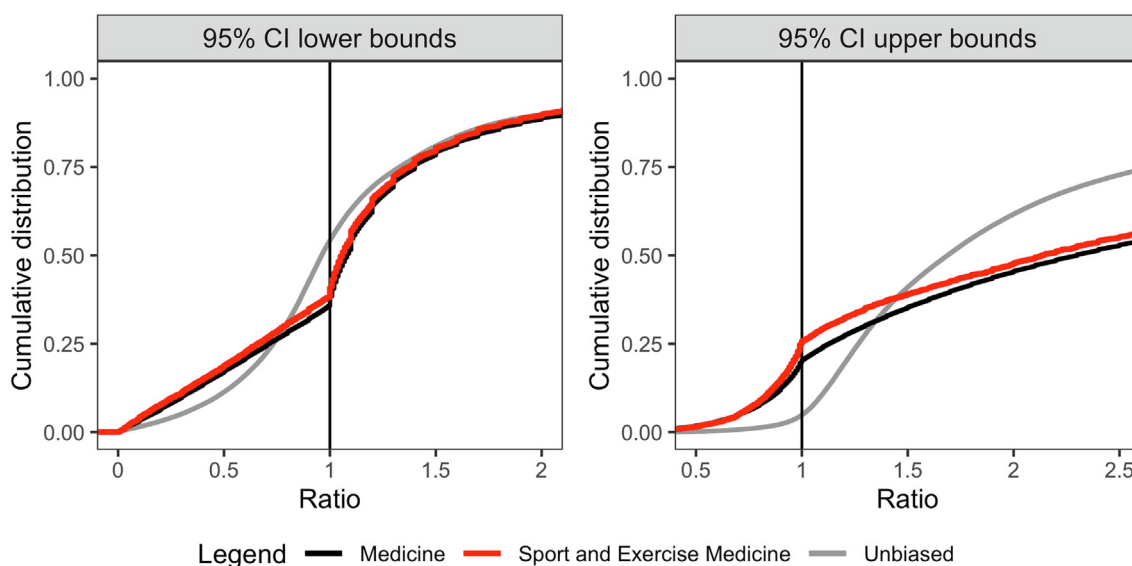


Fig. 1. Empirical cumulative distributions for ratio confidence intervals from the abstracts of articles published in sports and exercise medicine journals between 2002 and 2022 (red), the abstracts of articles published in medical journals between 1976 and 2019 (black), and from an unbiased reference dataset (gray). Lower bounds are shown on the left panel and upper bounds on the right panel. To be statistically significant, lower intervals need to be above 1, and upper intervals need to be below 1. The x-axes are restricted to focus on changes around the significance threshold of 1 (vertical line). Note the marked change in the distribution of intervals from sports and exercise medicine around a ratio of 1, which is not present in the distribution from the unbiased dataset. The marked change around a ratio of 1 was also evident for intervals from medicine. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

excess of results just below the significance threshold would not occur if published results were completely unbiased. Transparent research practices are needed to reduce the bias in published sport and exercise medicine research. This includes the use of registered reports,²⁵ ending the practice of continuing data collection until reaching significance, and the sharing of data and code. There is an urgent need for peer reviewers, editors, and journals to direct the rewards of publication away from statistical significance and onto scientific rigor.

Despite a smaller sample size, our findings in sports and exercise research are consistent with observations in medical research, where a large excess of lower and upper bound intervals around a ratio of 1 has been reported.¹³ We observed an abnormal change in the direction of the cumulative distribution around a ratio of 1, which is unlikely to occur in the absence of bias (Fig. 1). Similarities of the bias in confidence

intervals between medicine and sport and exercise medicine are further supported by the highly unusual distribution of z-values, characterized by a stark absence of non-significant z-values (Fig. 2), which was also observed in medicine (see Fig. 1 in van Zwet & Cator¹⁹).

Only focusing on statistically significant results is harmful for new discoveries because it distorts the literature by emphasizing an arbitrary threshold rather than rigor. Significant results with a small p -value are often mistakenly viewed as valid, reliable and meaningful,²⁶ yet the exclusive focus on significance exaggerates the magnitude of an effect.^{19,27,28} This bias decreases for larger sample sizes, which is worrying in the field of sport and exercise medicine, as sample sizes are often small, and therefore the bias is likely to be large.²⁹ The exaggeration of effects and subsequent distortion of evidence for scientific claims can lead to wasted resources, as researchers direct their attention toward

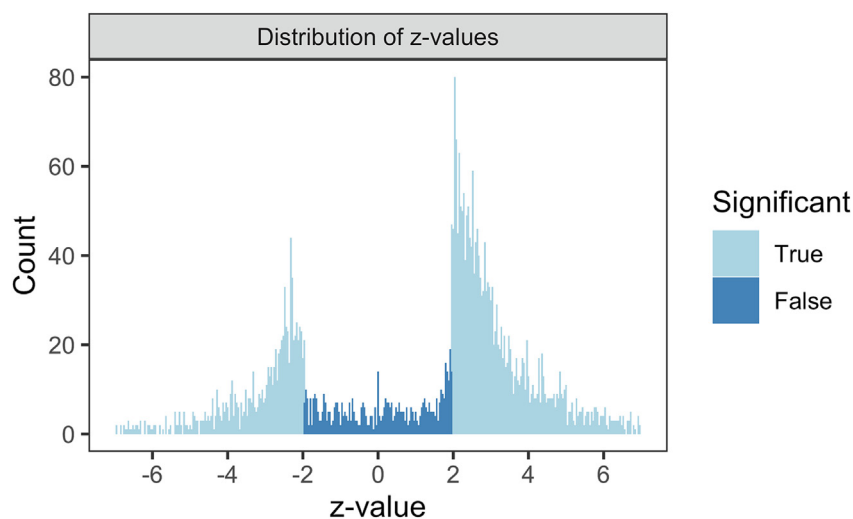


Fig. 2. The distribution of z-values. There was an under-representation of z-values between -1.96 and $+1.96$, corresponding to a p -value of 0.05, which is the commonly used significance threshold. The absence of published non-statistically significant results is striking. We do not know what the exact shape of the distribution would be in the absence of bias; however, we would expect the distribution to be flatter and smoother, with no spikes around z-values of -1.96 or $+1.96$. Note, histograms group data into "bins" to create a distribution. A user is required to specify a bin width, which depending on the choice, can create different impressions of the same data. We generated a high-resolution histogram using the bin width of 0.04, which provides a fair impression for our context.

unworthy areas, for which there is little evidence.^{8,7} Worse, unproven, or ineffective treatments may be promoted, which can directly harm the public and lower trust in scientific institutions.

If researchers focused on estimation, rather than significance, the exaggeration of effects could be reduced.³⁰ This would require researchers to think more carefully about their analysis and interpretation. Recently, there has been advocacy for adopting an unconditional interpretation of statistical results.^{30,31} This approach would involve focusing on the estimation of effects rather than statistical significance and focusing on the uncertainty around the estimated effect (e.g., the confidence interval width). It is believed that this unconditional estimation approach would avoid the problem of oversimplifying results into significance and non-significance.³⁰ However, there is no empirical evidence that shows requiring researchers to adopt such an approach reduces bias and improves the interpretation of statistical results. Bayesian estimation offers an alternative approach to relying on statistical significance, as effects can be directly quantified using posterior probabilities.^{32,33} Further, some argue that Bayesian credible intervals have a more direct and natural interpretation than frequentist confidence intervals, and the Bayesian approach does not rely on the same asymptotic assumptions that are often made to construct confidence intervals.³³ However, we note that thresholds can still be applied when using a Bayesian approach, so the approach is not a panacea for avoiding dichotomous thinking.

Improvements in research transparency are urgently needed. This includes pre-registration, the use of registered reports²⁵ and the public sharing of data and code. As a reminder, a registered report is a type of journal article where authors detail their study plan, including methods and analyses, which undergoes peer review and if passed the journal commits to publishing the results. In psychology, registered report studies produced far fewer positive results (44 %) than non-registered report studies (96 %), with “positive” being statistically significant.⁷ The success of registered reports in practice requires investment from several parties.³⁴ Reviewers and editors need to hold researchers to their pre-specified plan, allowing for reasonable exceptions due to unforeseen changes. To be effective, researchers must use registered reports. However, their use in the field remains sparse. For example, in the three years since *Science and Medicine in Football* introduced registered reports, the journal has received none.³⁵ This is not an isolated example. In the related field of sports science, only several registered reports have been published in the *Journal of Sports Sciences* since their introduction.³⁶ Registered reports are a long-term solution to improving research transparency, requiring systemic uptake by the field.

Journals have a critical role to play in research transparency, particularly through policy. For example, including the option for registered report submissions and mandating the sharing of sufficient data to replicate key results along with all computed generated code, with only rare exceptions. When the original data cannot be shared for privacy or other reasons, sharing a simulated data set based on the original, or a subset of data without identifying information are suitable alternatives. As an example, a subset of data could include physiological and performance data related to the experiment but not participant information. Data sharing is a skill and not all researchers have the capacity to simulate data.³⁷ In such cases, the use of existing technology to generate a synthetic dataset based on the original may be a user-friendly alternative.³⁷ We appreciate that researchers may have limited experience with open science and limited support for open science practices.³ Rather than making suggestions that place extra burden on researchers, we are advocating for data curation and sharing to become part of normal practice supported by research institutions. We hope that the establishment of the Society for Transparency, Openness, and Replication in Kinesiology (STORK) will improve research transparency in the field, including the widespread use of registered reports.

A limitation of the current study is that the text-mining algorithm does not capture a general statement about statistical significance, or lack thereof, when it is not accompanied by a confidence interval. We

examined confidence intervals reported in abstracts only. We acknowledge that abstract word limits may prevent the reporting of all key study outcomes. However, we do not believe that this would affect the substantive conclusion of our study. Previous work in medicine found that the distribution of lower and upper ratio confidence intervals reported in abstracts was similar to those reported in full-texts (see Fig. 1 in Barnett & Wren¹³). Additional analysis of articles published between 1976 and 2019 in journals indexed in *MEDLINE* containing the word “sport” or “exercise” in the title shows that the distribution of z-values calculated from ratio confidence intervals was similar between abstracts and full-texts (see Supplement 4). We summarized confidence intervals graphically and descriptively, rather than with any statistical model. About 11 % of the extracted interval pairs were missing the level of confidence (Supplement 2). Although we excluded these from the analysis, we found that the cumulative distribution of these 508 interval pairs was very similar to the main analysis, see Supplement 5. Our sample size was smaller than previous similar work.¹³ The relatively small sample size precluded some analyses, such as examining the cumulative distribution across journals, with some journals contributing less than 100 intervals to the data (Supplement 1). Nonetheless, our results clearly highlight the extent of bias in ratio confidence intervals in sport and exercise medicine, and can be considered robust given the similarity to observations in medicine where nearly 1 million ratio interval pairs were examined.¹³

5. Conclusion

There was an excess of published research with results that were just below the standard significance threshold of 0.05, which clearly shows publication bias. Transparent research practices are needed to reduce the bias in published sport and exercise medicine research, such as the use of registered reports and the sharing of study materials, including data and code. The successful implementation of registered reports in practice requires investment from authors and journal editors, and will need journal policy changes.

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jsams.2023.03.002>.

Funding information

No funding was awarded for this project.

Confirmation of ethical compliance

No ethical approval for the study was needed as we used publicly available data that is published to be read and scrutinized.

CRediT authorship contribution statement

Authors DNB, AGB and IBS designed the study. DNB and AGB extracted the data. DNB, AGB and ARC were all involved in the data analysis. All authors were involved in the interpretation of the data, drafting of the manuscript and subsequent revisions, and approved the final version of the manuscript before journal submission. All authors agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Data availability

The data and R code are available at <https://doi.org/10.5281/zenodo.7102324>. The code was adapted from <https://github.com/agbarnett/intervals>.

Declaration of interest statement

The authors have no competing interests to declare.

Acknowledgements

The authors would like to thank Dr Lucas BR Orssatto and Dr Robert Buhmann for their valued feedback.

References

- Buchanan TL, Lohse KR. Researchers' perceptions of statistical significance contribute to bias in health and exercise science. *Meas Phys Educ Exerc Sci* 2016;20(3):131-139.
- Emerson GB, Warne WJ, Wolf FM et al. Testing for the presence of positive-outcome bias in peer review: a randomized controlled trial. *Arch Intern Med* 2010;170(21):1934-1939.
- Twomey R, Yingling V, Warne J et al. The nature of our literature: a registered report on the positive result rate and reporting practices in kinesiology. *Commun Kinesiol* 2021;1(3):1-17.
- Fanelli D. How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PLoS One* 2009;4(5):e5738.
- Simonsohn U, Nelson LD, Simmons JP. p-Curve and effect size: correcting for publication bias using only significant results. *Perspect Psychol Sci* 2014;9(6):666-681.
- Sainani KL, Borg DN, Caldwell AR et al. Call to increase statistical collaboration in sports science, sport and exercise medicine and sports physiotherapy. *Br J Sports Med* 2021;55(2):118-122.
- Scheel AM, Schijen MR, Lakens D. An excess of positive results: comparing the standard Psychology literature with Registered Reports. *Adv Methods Pract Psychol Sci* 2021;4(2). 25152459211007468.
- Büttner F, Toomey E, McClean S et al. Are questionable research practices facilitating new discoveries in sport and exercise medicine? The proportion of supported hypotheses is implausibly high. *Br J Sports Med* 2020;54(22):1365-1371.
- Gelman A, Loken E. Gelman: the garden of forking paths: why multiple comparisons can be a problem, even when there is no "fishing expedition" or "p-hacking" and the research hypothesis was posited ahead of time. http://www.stat.columbia.edu/~gelman/research/unpublished/p_hacking.pdf 2013. Accessed 18 May 2022.
- Elkins MR, Pinto RZ, Verhagen A et al. Statistical inference through estimation: recommendations from the International Society of Physiotherapy Journal Editors. *J Physiother* 2022;68(1):1-4.
- Ranstam J. Why the P-value culture is bad and confidence intervals a better alternative. *Osteoarthr Cartil* 2012;20(8):805-808.
- Greenland S. Invited commentary: the need for cognitive science in methodology. *Am J Epidemiol* 2017;186(6):639-645.
- Barnett AG, Wren JD. Examination of CIs in health and medical journals from 1976 to 2019: an observational study. *BMJ Open* 2019;9(11):e032506.
- Hoekstra R, Morey RD, Rouder JN et al. Robust misinterpretation of confidence intervals. *Psychon Bull Rev* 2014;21(5):1157-1164.
- Morey RD, Hoekstra R, Rouder JN et al. The fallacy of placing confidence in confidence intervals. *Psychon Bull Rev* 2016;23(1):103-123.
- Georgescu C, Wren JD. Algorithmic identification of discrepancies between published ratios and their reported confidence intervals and P-values. *Bioinformatics* 2018;34(10):1758-1766.
- Szumilas M. Explaining odds ratios. *J Can Acad Child Adolesc Psychiatry* 2010;19(3):227.
- Schuemie MJ, Ryan PB, Hripcsak G et al. Improving reproducibility by using high-throughput observational studies with empirical calibration. *Philos Trans A Math Phys Eng Sci* 2018;376(2128):20170356.
- van Zwet EW, Cator EA. The significance filter, the winner's curse and the need to shrink. *Statistica Neerlandica* 2021;75(4):437-452.
- R Core Team. R: a language and environment for statistical computing. Version 4.1.3. <https://www.r-project.org> 2022.
- Wickham H, Averick M, Bryan J et al. Welcome to the Tidyverse. *J Open Source Softw* 2019;4(43):1686.
- Winter DJ. Rentrez: an R package for the NCBI EUtils API. *PeerJ Preprints* 2017:1-8.
- Temple LD. XML: tools for parsing and generating XML within R and S-Plus. R package version 3.99-0.9: <https://CRAN.R-project.org/package=XML> 2022.
- Tierney N, Cook D, McBain M et al. naniar: Data structures, summaries, and visualisations for missing data. R package version 0.6.1: <https://CRAN.R-project.org/package=naniar> 2021.
- Caldwell AR, Vigotsky AD, Tenan MS et al. Moving sport and exercise science forward: a call for the adoption of more transparent research practices. *Sports Med* 2020;50(3):449-459.
- Berner D, Amrhein V. Why and how we should join the shift from significance testing to estimation. *J Evol Biol* 2022;35:777-787.
- van Zwet E, Schwab S, Greenland S. Addressing exaggeration of effects from single RCTs. *Significance* 2021;18(6):16-21.
- Colquhoun D. An investigation of the false discovery rate and the misinterpretation of p-values. *R Soc Open Sci* 2014;1(3):140216.
- Hutchins KP, Borg DN, Bach AJ et al. Female (under) representation in exercise thermoregulation research. *Sports Med Open* 2021;7(1):1-9.
- Rafi Z, Greenland S. Semantic and cognitive tools to aid statistical science: replace confidence and significance by compatibility and surprise. *BMC Med Res Methodol* 2020;20(1):1-13.
- McShane BB, Gal D, Gelman A et al. Abandon statistical significance. *Am Stat* 2019;73(1):235-245.
- Borg DN, Minett GM, Stewart IB et al. Bayesian methods might solve the problems with magnitude-based inference. *Med Sci Sports Exerc* 2018;50(12):2609-2610.
- Kruschke JK, Liddell TM. The Bayesian New Statistics: hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychon Bull Rev* 2018;25:178-206.
- Scheel AM. Registered reports: a process to safeguard high-quality evidence. *Qual Life Res* 2020;29(12):3181-3182.
- Impellizzeri FM, McCall A, Meyer T. Registered reports coming soon: our contribution to better science in football research. *Sci Med Footb* 2019;3(2):87-88.
- Abt G, Boreham C, Davison G et al. Registered reports in the journal of sports sciences. *J Sports Sci* 2021;39(16):1789-1790.
- Quintana DS. A synthetic dataset primer for the biobehavioural sciences to promote reproducibility and hypothesis generation. *eLife* 2020;9:e53275.