Title

Overcoming the “Ostrich Effect”: A Narrative Review on the Incentives and Consequences of Questionable Research Practices in Kinesiology

Journal

Kinesiology Review, 12(3)

ISSN

2163-0453

Authors

Tiller, Nicholas B
Ekkekakis, Panteleimon

Publication Date

2023-08-01

DOI

10.1123/kr.2022-0039

Peer reviewed

Overcoming the "ostrich effect": A narrative review on the incentives and consequences of questionable research practices in kinesiology

Nicholas B. Tiller[1], Panteleimon Ekkekakis[2]

POSTPRINT: 25 Feb 2023

[1]Institute of Respiratory Medicine and Exercise Physiology, The Lundquist Institute for Biomedical Innovation at Harbor-UCLA Medical Center, Torrance, CA, USA.

[2]Department of Kinesiology, Michigan State University, East Lansing, MI, USA.

Correspondence

Nicholas B. Tiller, Ph.D.

1124 W. Carson Street, CDCRC Building, Torrance, CA 90502.

Email: nicholas.tiller@lundquist.org | Orchid: https://orcid.org/0000-0001-8429-658X

Running head: Questionable research practices in kinesiology

Citation: Tiller NB, Ekkekakis P. (2023). Overcoming the "ostrich effect": A narrative review on the incentives and consequences of questionable research practices in kinesiology. *Kinesiology Review, EPub ahead of print*

Abstract

Increasing transparency and openness in science is an ongoing endeavor, one that has stimulated self-reflection and reform in many fields. However, kinesiology and its related disciplines are among those exhibiting an “ostrich effect” and a reluctance to acknowledge their methodological shortcomings. Notwithstanding several high-profile cases of scientific misconduct, scholars in the field are frequently engaged in questionable research practices (QRPs), such as biased experimental designs, inappropriate statistics, and dishonest/inexplicit reporting. To advance their careers, researchers are also “gaming the system” by manipulating citation metrics and publishing in predatory and/or pay-to-publish journals that lack robust peer-review. The consequences of QRPs in the discipline may be profound: from increasing the false positivity rate to eroding public trust in the very institutions tasked with informing public-health policy. But what are the incentives underpinning misconduct and QRPs? And what are the solutions? This narrative review is a consciousness raiser that explores i) the manifestations of QRPs in kinesiology; ii) the excessive publication pressures, funding pressures, and performance incentives that are likely responsible; and iii) possible solutions for reform.

Key words: ethics; exercise; health; methods; science; statistics.

1. Introduction

In one of the largest scientific misconduct cases on record, an American physiologist admitted falsifying data in 10 published articles and 17 grant applications worth nearly $3 million (USD) (Dahlberg & Mahler, 2006; Dalton, 2005; Kondro, 2005; Sox & Rennie, 2006). At the conclusion of his federal trial, Eric Poehlman was sentenced to one year (and one day) in prison for submitting fraudulent data on the science of obesity, menopause, and aging, heralding the first time a scientist in the United States had been jailed for research misconduct that did not result in fatalities (Kintisch, 2006). In a letter to the judge, Poehlman asked for leniency and confessed: "I was motivated by my own desire to advance as a respected scientist" (Kintisch, 2006).

Misconduct also manifests at the corporate level, affecting multinational businesses and even sports medicine agencies (Serodio et al., 2020). For instance, when concerns about obesity triggered a decline in the consumption of sugar-sweetened beverages, the Coca-Cola Company, via The Global Energy Balance Network, donated at least $1.5 million to research that would ultimately downplay the role of poor diet and calorie control in weight management (Walters, 2015). Not only did such a view “fall outside the scientific consensus” (Krans, 2022), but studies revealed substantial conflicts of interest and reporting bias in the subsequent literature (Barlow et al., 2018; Bes-Rastrollo et al., 2013; Stuckler et al., 2018).

A phenomenon more subtle and deep-rooted in institutional norms than misconduct, and which may do more long-term harm to scientific enquiry, is questionable research practice (QRP) (John et al., 2012; R. Schulz et al., 2022). Questionable research practices are introduced deliberately or inadvertently into study design (e.g., biased and poorly controlled experiments), data collection (e.g., insufficient blinding), data analysis (e.g., incorrect or inappropriate statistical procedures), and data reporting (e.g., *post hoc* hypothesizing), leading to non-replicable results and conclusions (Büttner et al., 2020). By enabling scientists to manipulate the results of their research, QRPs create a system in which honest researchers are at a competitive disadvantage. For this reason, QRPs have been described as “the steroids of scientific competition” (John et al., 2012). The self-confessed prevalence of QRPs across scientific disciplines has been reported as high as 51% (Fanelli, 2009; Gopalakrishna et al., 2022; John et al., 2012), with rates rising to 72% when scientists were asked about the QRPs of their colleagues (Fanelli, 2009).

Research misconduct and QRPs in kinesiology (herein denoting all its related disciplines including sports medicine; physical education; and the sports, health, and exercise sciences) may have broad consequences: from increasing the frequency of false positives in the published literature, to diminishing scientific quality and rigor, and inhibiting scientific progress and the attainment of replicable scientific knowledge. To a large degree, ambiguity and lack of transparency in some kinesiology research is being exploited by the commercial health and wellness industries to sell products and practices on baseless claims and pseudoscience (Tiller, 2020; Tiller et al., 2022). Misconduct may also be damaging the reputation of the discipline (Tiller et al., 2022) and, therefore, harming graduate employment prospects (Yong, 2012).

The problem of research misconduct and QRPs in kinesiology has been the subject of much discussion (Atkinson & Nevill, 2001; Caldwell et al., 2020; Earnest et al., 2018; Halperin et al., 2018; Knudson, 2012, 2017a; Marticorena et al., 2021;

Twomey et al., 2021). Although the prevalence of QRPs in kinesiology has not been directly studied, the number of article retractions suggests the rate of QRPs in the discipline may be increasing. One systematic review reported that 52 articles had been retracted from “sports science” journals between 1979 and 2018, with more than half of the retractions (*n*=28) occurring in the last decade (Kardeş et al., 2020), and with most being attributed to misconduct (44%) rather than honest error (37%). The analysis also showed a slight increase in the *rate* of retracted papers in “sports science” when expressed relative to the total number of published papers between 2000 and 2018 (Kardeş et al., 2020). These numbers are approximately comparable to data from other disciplines, including the fields of biomedical (Gasparyan et al., 2014), surgical (King et al., 2018), and intensive-care (Wiedermann, 2016) medicine.

Aside from isolated editorials and opinion pieces, the issues of open science and replicability in kinesiology have not been discussed in edited volumes, journal special issues, or conference symposia. The apparent lack of attention to these issues may indicate a striking lack of awareness, and possibly lack of concern, regarding the discipline’s widespread and arguably severe methodological shortcomings (Atkinson & Nevill, 2001; Sainani et al., 2021). If QRPs continue unabated and unaddressed, and their prevalence exceeds a critical threshold, the institution of kinesiology may be in danger of becoming irreversibly damaged. This will slow the attainment of new knowledge and further erode public trust in the very institutions tasked with informing public health policy. The consequences for humanity would be profound (Edwards & Roy, 2016).

This narrative review is a consciousness raiser for kinesiology academics and practitioners. Herein, we provide discipline-specific examples of the most common

QRPs in kinesiology, and explore the institutionalized, quantitative academic-performance incentives that are likely responsible. We conclude by providing recommendations for reforming the current knowledge dissemination paradigm. With this call to action, we hope to stimulate open dialogue among academics and practitioners on how we can progressively mitigate research misconduct and QRPs in the discipline. Indeed, transparent communication is the crucial first step on a long journey toward broader awareness and meaningful change.

2. The Incentives Underpinning Research Misconduct

Scientific misconduct and QRPs persist across scientific fields despite ongoing education and repeated calls for change. This suggests that their existence is less attributable to procedural misunderstandings and more attributable to incentives (Edwards & Roy, 2016). There is a growing body of evidence that these incentives relate to professional ambitions (John et al., 2012), academia’s fixation on quantitative productivity metrics (Biagioli & Lippman, 2020; Edwards & Roy, 2016), the hypercompetitive funding environment (Martin, 2020), the changing business model of higher education (Edwards & Roy, 2016; Gerrits et al., 2019; Van Noorden, 2010), and financial inducements (Lesser et al., 2007; Lucas, 2015).

Publication pressure, and the increased competition to publish impactful research, is at the center of this multifaceted incentive structure. A recent bibliometric analysis showed that, over the past several decades, the number of published scientific papers has increased by 8–9% each year, representing more than 1 million new papers in the biomedical field alone—approximately two new publications every minute (Landhuis, 2016). There is also an increased appetite to publish exercise-related research. To illustrate this, we selected “sports science” as the most closely related category listed in Scopus® (the largest curated database) and analyzed publication data for all subdisciplines and regions/countries between 2000 and 2021 (**Figure 1**). The total number of published articles increased 2.4-fold, from 7,655 to 18,564, with a steady increase in the number of submissions to the top 50 highest-ranked journals (listed in ascending order of the average number of weighted citations for a given year relative to the documents [*n*] published in the journal in the three previous years), from 6,403 to 10,817. The total number of journals also increased from 83 to 121 between 2000 and 2011 and plateaued thereafter **(Figure 1)**. In a published analysis, the number of articles submitted to just one section of the *Journal of Sports Sciences* between 2017 and 2020 increased by 34% (from 637 to 854) (Abt et al., 2022). This increase in the journal's popularity is presumed to be attributable to increased publication pressures (Brischoux & Angelier, 2015), increased data availability through routine monitoring of athletes (Robertson, 2020), and perverse incentives and metrification driving academics and researchers to seek high volume output (Edwards & Roy, 2016; Knudson, 2019). Publishing in the health sciences has become "almost compulsory…" (Dinis-Oliveira & Magalhães, 2016), and we assert that this also applies to kinesiology more broadly.

Pertinently, evidence suggests that publication pressures are independently contributing to the prevalence of QRPs. In a survey of nearly 7,000 scientists across various disciplines, "publication pressures" associated positively with the frequency of QRPs, as did being a doctoral candidate, a junior researcher, and being male (Gopalakrishna et al., 2022). In a smaller survey of ~600 biomedical scientists, "publication pressure" emerged as the strongest individual predictor of misconduct ($\beta = 0.34$, $p < 0.01$), accounting for 10% of the variance in the outcome (Maggio et al., 2019). In addition, younger researchers reported more misconduct—perhaps due to greater promotion pressures, tenure pressures, and/or less familiarity with responsible research practices (Fanelli et al., 2015; Maggio et al., 2019). A third study found that publication pressures among medical scientists associated significantly ($\beta = 0.07$, $p < 0.001$) with a composite misconduct-severity score, both in univariate analyses and after adjustment for demographics (Tijdink et al., 2014). Although the factors linking publication pressures and QRPs are not definitively known, Pabst *et al*. (2013) showed that psychological stress in gain-and-loss tasks increased the frequency of risky decisions. Tijdink *et al*. (2014) therefore hypothesized that stress from the demand to publish may increase risky behavior in research, contributing to scientific misconduct.

It has been proposed that QRPs may be necessary just to “survive in academia” (van de Schoot et al., 2021). While it could be argued that quantitative performance metrics facilitate healthy competition that, in turn, drives scientific progress, these same metrics are likely a poor indicator of socially relevant and impactful research (Biagioli & Lippman, 2020; Schmid, 2017). It is also likely that academia’s never-ending selection for productivity in research—reflected by pervasive idioms like “publish or perish” (Brischoux & Angelier, 2015)—has led to a preference for research quantity over quality. In fact, an analysis of the 100 top-cited articles in “sports science” and “sports medicine” (extracted from Web of Science, Scopus, and PubMed until 2019) returned 38 narrative reviews and only one randomized controlled trial, all of which were published between 1973 and 2013 (Khatra et al., 2021).

Universities and journals are equally implicated in popularizing the perverse culture of “publish or perish”. Facing budgetary pressures, academic institutions rely on prestige (gained partly through their visibility in high-profile journals) to attract research funding. The resulting pressure on researchers to increase the frequency of manuscript submissions (Moylan & Kowalczuk, 2016) subsequently elevates journal operating costs. With the number of “sports science” journals increasing by ~45% since 2000 (see **Figure 1**), superiority in this saturated space is attained via widely advertised impact factors and other measures of perceived rank that create a preference for findings that are most likely to yield citations and media attention. The proclivity to produce such findings is then, unsurprisingly, reflected in the research practices of career scientists. The current incentive structure, therefore, encourages research misconduct and QRPs by influencing the priorities of journals, institutions, and researchers in a perpetual and reciprocal manner.

It is also likely that publication pressures underpin the increasing popularity of predatory journals. Predatory journals are appearing at a rate that exceeds “reliable” journals (The Economist, 2020). An analysis by Shen & Björk (2015) showed that ~53,000 articles were published in predatory journals in 2010, increasing to ~420,000 articles in 2014. The average time from submission to publication in such journals was 2.7 months (considerably faster than most mainstream outlets) with a mean publication fee of just $178 (Shen & Björk, 2015). Pertinently, the authors believe that most researchers are not victims of predatory journals but are instead aware of the circumstances surrounding publication and are making “calculated risks that experts who evaluate their publication lists…”, such as academic search, award, or promotion-and-tenure committees, “…will not bother to check the journal credentials in detail” (Shen & Björk, 2015). Most articles (~60%) published in predatory journals receive few, if any, citations in the five years after publication, supporting the notion that researchers utilize these outlets to inflate publication number rather than citations metrics (Brainard, 2020).

Financial incentives have also been shown to drive publication pressures and bias study conclusions, specifically in nutrition-related research (Lucas, 2015). For instance, the worldwide sports supplement industry is worth an estimated $40 billion (Statista, 2020). Since the *Nutrition and Health Claims Regulation of 2012*—which required supplement manufacturers to produce evidence-for-efficacy from human studies—there has been a considerable proliferation of sports supplement research (Kiss et al., 2021), much of it industry funded. Publishing highly marketable nutrition research could be commercially profitable, but several independent analyses revealed that industry-funded nutrition research is more likely to report favorable outcomes when compared to non-industry-funded research (Bes-Rastrollo et al., 2013; Diels et al., 2011; Lesser et al., 2007). In turn, this may "bias conclusions in favor of sponsored products, with potentially significant implications for public health" (Lucas, 2015). Those working in kinesiology are also facing mounting pressure from funding agencies to bridge basic and translational studies (Sabroe et al., 2007) and minimize the gap between the laboratory and the playing field (Eisenmann, 2017). However, by competing to address (and solve) real-world problems through scientific exploration, researchers may be overgeneralizing results and prioritizing marketable research instead of emphasizing scientific rigor, resulting in low-quality studies, biased interpretations, and inconsistent reporting in areas that include sports equipment (Bachynski & Smoliga, 2021; Smoliga, 2020) and training programs (Ekkekakis & Tiller, 2022).

Lastly, there are more direct examples of financial incentives impinging on transparency in research. Until 2020, numerous academic institutions, predominantly in China but also in the West, offered academics cash rewards for publishing in journals that were indexed by Web of Science and that surpassed minimum impact factors. The more prestigious the outlet, the higher the reward, with a manuscript in *Science* or *Nature* worth an average of $43,000 (Quan et al., 2017). Such a system is thought to promote "perverse incentives" that prioritize productivity over rigor (Mallapaty, 2020).

In summary, many factors denoting scientific "fitness" in contemporary academic culture—publication numbers, grant income, quantitative productivity or impact metrics, marketable research—are the same factors that have been identified as possible culprits in incentivizing misconduct and QRPs. Not only does the system benefit researchers that produce high-volume, low-quality output, but it may also be disaffecting academics with the strongest ethical and moral principles. The result is the "natural selection" of bad science (Smaldino & McElreath, 2016). Data show that doctoral graduates in the biomedical sciences tend to pursue careers that align with their core beliefs (Gibbs & Griffin, 2013). Accordingly, a continued emphasis on performance metrics over altruistic values may risk alienating the next generation of researchers in the Science, Technology, Engineering, and Math (STEM) field. These probable outcomes are antithetical to the broader ambitions of scientific integrity and may have irreversible repercussions for kinesiology and related disciplines.

3. Examples of Misconduct and Questionable Practices

Misconduct and the full range of QRPs manifest in kinesiology research because of perceived flexibility in study design, data collection, statistical analysis, and interpretation or reporting. Rather than provide an extensive overview of each

QRP, the focus here will be on those that manifest most prominently in kinesiology and for which there are readily available examples in the literature.

Publication bias. The notion that researchers are less likely to submit, and journals less likely to accept, papers that show negative or neutral outcomes, is not a new one. In fact, there is a significant positive relationship between study outcomes and a researcher's decision to submit a paper for review (Coursol & Wagner, 1986). This phenomenon, colloquially referred to as 'the file drawer problem', is a challenge to kinesiology research (Bernards et al., 2017; The Society for Transparency, Openness, and Replication in Kinesiology, n.d.), leading to publication bias in the exercise (Twomey et al., 2021), biomedical (Fanelli, 2010; Sterling et al., 1995), psychological (Scheel et al., 2020), and social sciences (Franco et al., 2014). In sports medicine and related research, only 59% of pre-registered trials were eventually published (Chahal et al., 2012), and those that were published exhibited discrepancies between the published article and the registered protocol in at least one methodological element (primary/secondary outcomes, inclusion/exclusion criteria, sample size; Chahal et al., 2012). In kinesiology, the rate of "positive" results has been reported as ~81% (Twomey et al., 2021). An analysis of sport and exercise psychology research showed that ~98% of studies reported at least one significant finding, with ~80% rejecting the main stated null hypothesis (Spence & Blanchard, 2001). This indicates that publication bias is "alive and well in the sport and exercise psychology literature" (Spence & Blanchard, 2001, p. 386). Accordingly, the data indicate a strong and systematic bias toward study designs and methodological choices that tend to confirm researcher expectations (Büttner et al., 2020). Publication bias hinders scientific progress because researchers squander time and resources exploring effects that may not be valid or replicable, particularly given that nonreplicable studies are cited more often than replicable ones (Serra-Garcia & Gneezy, 2021). The imperative to conduct more replication studies in kinesiology and related disciplines has been discussed (Knudson, 2017b). Moreover, a large collaborative effort is underway to assess replicability in sports and exercise science research (Murphy et al., 2023), in the same vein that other projects have evaluated the replicability of research in psychology (Open Science Collaboration, 2015) and cancer biology (Errington et al., 2021).

Confirmatory versus exploratory studies. Kinesiology is a relatively young field with grounding principles and operating standards less well established when compared to sciences with deeper historical roots. As a result, the majority of studies published in related journals are exploratory/observational rather than confirmatory (Bleakley & MacAuley, 2002; Twomey et al., 2021). In other words, for most kinesiology research, the full and transparent registration of aims, hypotheses, methodologies, and statistical analyses prior to data collection would not have been possible (Bleakley & MacAuley, 2002; Harris et al., 2014). But when exploratory research is falsely reported as confirmatory, there is an increased risk of inaccurate, erroneous, or non-replicable outcomes (Büttner et al., 2020; Ioannidis, 2005). Not only does this further obscure the interpretation of study findings, it also contributes to inflated false positives and is partly responsible for the replication crisis (Begley & Ioannidis, 2015; Nosek, 2015). We emphasize that there is nothing inherently wrong with conducting exploratory studies or with inductive reasoning: the problem lies in the misrepresentation of exploratory studies as confirmatory. Significant results, especially those that are statistically robust and derived from properly conducted and analyzed exploratory studies, should be transparently reported as such and followed-up with rigorously designed and sufficiently powered confirmatory studies.

Post hoc hypotheses. The anticipated outcomes of interventional studies are typically stated in advance. However, sometimes a hypothesis is generated retroactively (only after the data have been analyzed) but then presented in the manuscript as though designed *a priori*. Such *post hoc* hypothesizing has been referred to as "hypothesizing after the results are known" (*HARKing*), and may be attributable to a poor understanding of research practice as opposed to deliberate deception (Kerr, 1998). One analysis of sport and exercise medicine research revealed that only 60% of published studies stated an *a priori* hypothesis and, of those that did, 82% reported findings that supposedly confirmed the hypothesis (Büttner et al., 2020) (Table 1). Given that *HARKing* can occur in the context of interventional or exploratory research that may lack a directional hypothesis, the practice bypasses an important safeguard against Type I errors (false positives), which is inherent in all statistical analyses.

Self-citations. Because research is a continuous and systematic process, with each investigation building on the last, a certain degree of self-citation is inevitable and even necessary. However, because quantitative metrics like citation number and H-index (a measure of the number of publications for which an author has been cited at least that same number of times) are used as indicators of research impact and excellence (Hicks et al., 2015), they are often abused. Using Scopus data, Ioannidis *et al*. (2020) published a database of citation metrics for 195,605 of the top-cited scientists (1960-2021) across 22 scientific fields and 176 sub-fields. We performed a sub-analysis on the 554 scientists for whom "Sport Sciences" was listed as the primary field of research (**Figure 2**). The mean (± SD) number of published articles among this subgroup was 177 ± 111 (range 22 to 754). The mean number of citations including self-citations was 7,752 ± 6,140 (range 1,121 to 42,416) and the mean number of citations excluding self-citations was 6,702 ± 5,381 (range 1,021 to 36,179). Pertinently, while the mean percentage of self-citations was 14 ± 7% (range 0% to 54%), six authors exhibited rates that exceeded three standard deviations from the mean: 54%, 44%, 42%, 41%, 37%, and 35%. In the database as a whole (all disciplines, n = 195,605 records [Ioannidis et al., 2020]), the mean self-citation rate was 13 ± 9%. Moreover, 825 scientists from various fields exhibited self-citation rates ≥ 50%, showing that a high rate of self-citation is not a phenomenon exclusive to “Sports Sciences”. Nevertheless, these data illustrate a clear propensity for a minority of “Sports Sciences” researchers to self-cite at a rate that is both extreme and anomalous according to the empirical rule. Although prolific self-citation is not necessarily unethical, it does highlight broader concerns about how researchers perceive citations and other quantitative metrics to influence hiring, promotions, pay, and research funding (Van Noorden & Singh Chawla, 2019).

Data fabrication/falsification. A study into the research practices of over 2,000 scientists at major U.S. universities reported data falsification prevalence estimates of 9% (John et al., 2012), although this is probably an underestimation of the true value. Besides a few high-profile misconduct cases like those of Eric Poehlman (Dahlberg & Mahler, 2006; Sox & Rennie, 2006) and Milena Penkowa (Callaway, 2011), there are no direct data on fabrication/falsification in kinesiology. However, a narrative review of misconduct in sports science research proposed that data fabrication was one of the primary manifestations of “abusive behavior” (Gaspar & Esteves, 2021), and there is little reason to think that kinesiology more broadly is an exception to trends in other fields. Further research to elucidate the extent of the problem in kinesiology is warranted.

4. Examples of Misconduct and Questionable Practices in Statistical Analyses and Reporting

There have been several calls for more robust and transparent statistical reporting in fields such as sports medicine (Altman et al., 1983; Gardner et al., 1983), biomechanics (Knudson, 2009), physiology (Curran-Everett & Benos, 2007), and psychology (Thompson, 1996). Multiple sets of author guidelines have also been published by *the American Psychological Association* (Cumming et al., 2012) among others. Nevertheless, the standards of statistical reporting have remained essentially unchanged (Diong et al., 2018a; Gandevia, 2021; Vagenas et al., 2018). From a statistical standpoint, research findings are more likely to be erroneous when studies are small, when effect sizes are small, and when there is a greater number (and lesser preselection) of tested relationships or effects (Ioannidis, 2005). The phenomenon of small samples and low statistical power in physiological, kinesiological, and psychological research was identified in the 1970s and described as a “faulty sampling practice” (Christensen & Christensen, 1977)—yet, it could be argued that these characteristics still define much of the present-day research in kinesiology and related disciplines. Here, we draw attention to these issues and to several other QRPs in statistical analyses and reporting that manifest frequently in the literature.

4.1 Over-reliance on p-values. It is traditional in statistical analyses for a *p*-value <0.05 to denote statistical significance. Since originating from the work of Cambridge geneticist and statistician R.A. Fisher in the 1920s (Fisher, 1926), this arbitrary threshold has been applied liberally and indiscriminately. For instance, a recent RCT in adults with cognitive impairment reported that exercise significantly improved executive function ($p = 0.046$) whereas a dietary intervention did not ($p = 0.059$) (Blumenthal et al., 2019), despite nearly identical pre- to post-intervention effect sizes. Given that there was no between-group difference in 'effectiveness', and likely no difference in clinical value, the example illustrates how an arbitrary *p*-value of 0.05 can obscure data interpretation.

An analysis of 300 original research articles from flagship kinesiology journals in North America (*Medicine and Science in Sports and Exercise*), Europe (*European Journal of Sport Science*), and Australia (*Journal of Science and Medicine in Sport*) showed that 92% of published studies relied on significance testing, 82% of which did not state an *a priori* hypothesis (Twomey et al., 2021). Moreover, simply obtaining a statistical test yielding a *p*-value < 0.05 provides little assurance that the result is replicable. In actuality, the odds of successfully replicating a significant result increase congruent with the *p*-value, and the relation is not what most researchers probably assume. Indeed, the odds are higher than 90% for *p*-values < 0.001, but drop to about 66% at a *p*-value of 0.01, and to 50% (i.e., chance) for *p*-values close to 0.05 (Curran-Everett, 2016; Goodman, 1992) (**Figure 3**). To improve the overall standards of reporting in kinesiology, those studies reporting *p*-values at the exclusion of all other descriptive and inferential statistics should be subjected to greater scrutiny. A 2014 meeting of the *American Statistical Association* highlighted a worrisome circularity in the use of the 0.05 statistical threshold: "We teach it because it's what we do; we do it because it's what we teach" (Wasserstein & Lazar, 2016).

A manifestation of the overreliance on significance testing is the use of *p*-values between 0.050 and 0.100 to denote output that is “approaching significance” or “trending toward significance”. This approach has been criticized because it is a subjective interpretation and because there is no category whereby one can “almost reject” the null hypothesis (Gibbs & Gibbs, 2015). Others describe it as “special pleading whereby authors, however unwittingly, are claiming something that their study has not achieved” (Wood et al., 2014). Researchers may also highlight “statistical trends” in studies they perceive to be underpowered due to low sample size; however, this is both misleading and quantitatively false. In fact, an analysis by Wood *et al*. (2014) showed that collecting more data on the premise that a study is 'underpowered' will result in *p*-values often getting larger, not smaller: e.g., collecting 10% more data will result in a “marginally non-significant” *p*-value of 0.08 getting smaller only 39% of the time. This underscores the instability or volatility (i.e., wide confidence intervals) of estimates derived from small samples.

4.2 Not correcting for inflation of familywise error rate. Conducting two independent statistical tests and evaluating each using the criterion of $p < 0.05$ can inflate alpha (the probability of committing a Type I error) to ~10%. Likewise, using the Šidák formula, one can estimate that conducting six independent tests, each using $p < .05$ to determine statistical significance, raises the likelihood of committing a Type I error to 26.49%. It only requires 14 independent tests for the risk of Type I error to surpass 50%. In other words, the likelihood of false positives increases along with the number of independent tests that are performed, hence the importance of adjusting the statistical output or the alpha level to account for multiple comparisons. An analysis of 232 studies from the field of "sports sciences" revealed a median of 30 statistical tests, while only 14% of them had specified a primary outcome (Lohse et al., 2020; Sainani & Chamari, 2022). Similar issues on a larger scale plague the field of genetic association research, which includes a growing number of studies pertaining to exercise and physical activity (e.g., Klimentidis et al., 2018; Williams et al., 2021). The field has been criticized for using insufficiently conservative statistics and capitalizing on “chance” to grossly exaggerate the extent to which genetic variants associate with the risk of disease and various health-related traits (Ioannidis et al., 2001; Prom-Wormley et al., 2017; Watanabe, 2011). Indeed, the number of genetic markers typically assayed in such studies can exceed 100,000, and many putative associations may occur by chance, even when using a seemingly "conservative" significance threshold of $p < 0.001$ (Teo, 2008).

4.3 Omitting effect sizes. Effect sizes have been described as “the most important outcome of empirical studies” (Lakens, 2013). This is for several reasons: (i) they enable researchers to quantify the magnitude of an effect and its practical significance using standardized criteria; (ii) they allow researchers to compare standardized effects among studies; and (iii) they facilitate the evidence synthesis and *a priori* power calculations for future studies (Lakens, 2013). The importance of effect size as a means of “describing the meaningfulness of findings” and protecting against misleading statistics in exercise-related research was discussed in the early 1990s (Thomas et al., 1991). Yet, evidence on the frequency of effect size reporting in kinesiology research remains mixed. An analysis of sports nutrition research showed that only 29% of studies reported effect sizes (Earnest et al., 2018) and this, in turn, entails an overdependence on *p*-values for interpreting results. Such infrequent reporting is considerably lower than the 81% of sport psychology studies that apparently show effect sizes (Andersen et al., 2007). By contrast, others have shown that "some form of effect size" was reported in sports-science journals at a rate of around 80% (Twomey et al., 2021). The most recent CONSORT statement (the "minimum" set of recommendations for reporting randomized trials) states that "For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval)", should be reported (Schulz et al., 2010). The publication manual of the *American Psychological Association* strongly also advocates the reporting of effect sizes (American Psychological Association, 2020), and it appears necessary to issue a similar mandate in kinesiology journals.

4.4 Not reporting/mis-reporting variance. There are widespread inconsistencies in the reporting of variance in sports nutrition/physiology research, with studies utilizing various combinations of standard deviation (SD), standard error of the mean (SEM), and confidence intervals (CI), sometimes interchangeably (Earnest et al., 2018). In premier physiology journals, ~80% of papers reported SEM as an estimate of variability (Diong et al., 2018), perhaps to conceal large variability in the data and subsequent plots. But SEM is not a measure of variability, rather it is a measure of uncertainty (Gandevia, 2021). Moreover, it should not be used as a descriptive statistic, but rather as an inferential one (Hopkins et al., 2009; Nagele, 2003). Indeed, the co-reporting of SD and CI is preferred over SEM (Hopkins et al., 2009). Another reason to include SD in original research (where relevant) is that it allows for the calculation of effect size (see section 4.3) and enables a given study to be included in future meta-analyses. Greater emphasis on distinguishing SD and SEM will improve the overall standards of reporting in kinesiology research.

4.5 P-hacking. When researchers explore numerous dependent measures and data processing/analytical approaches, and then report the outcome that provides the most novel, convenient, or intriguing results (or the ones that reach the threshold of $p < 0.05$), the process can be described as *p*-hacking (Silberzahn et al., 2018; Simmons et al., 2011). Of course, such flexibility in analytical procedures and reporting elicits multiple different outcomes using the same original data set, thereby increasing the likelihood of false positives (Simmons et al., 2011). Given that kinesiology research is rarely pre-registered, it is more likely that researchers will attempt multiple statistical analyses and then report the ones that best fit their hypotheses or biases (Caldwell et al., 2020). This underscores the importance of registering the planned analyses in advance of data collection. During interventional exercise studies, a more subtle form of *p*-hacking is to divide samples into “responders” and “non-responders” when, in fact, the superficial variability can often be explained by random within-subject day-to-day variation (Atkinson et al., 2015; Atkinson & Batterham, 2015). In fact, in physiological studies (Atkinson et al., 2019), in supplement studies (Del Coso et al., 2019), and in exercise-training studies (Montero & Lundby, 2017), the dichotomization of individuals into responders and non-responders has been criticized as fraught with pitfalls. Accordingly, while research on response variability is still needed, in part to determine if distinct categories of exercise responses exist, better care is necessary to distinguish a reproducible response from that evoked by random noise (Islam & Gurd, 2020; Padilla et al., 2021).

4.6 Sample size and statistical power. Calculating the minimum sample size for a study ensures adequate statistical power to detect an effect when one exists (i.e., when the null hypothesis is false). A sample too small will yield poor statistical power and imprecise population estimates, leading to inconclusive and non-replicable results (Vankov et al., 2014), whereas a sample too large will be financially costly and ethically questionable owing to unnecessary risks or inconveniences imposed on the participants. A power analysis is a solution that enables researchers to calculate *a priori* how many participants should be recruited in order to reduce the risk of errors of statistical inference (Jones et al., 2003). However, a minority of exercise-related studies utilize this important tool. An analysis of 120 randomly selected papers published in the *Journal of Sports Sciences* revealed that only 11% provided any formal *a priori* estimation of sample size (Abt et al., 2020). A separate analysis showed that sample sizes were appropriately justified in only 19 - 35% of studies published in kinesiology journals globally (Twomey et al., 2021). Although *Medicine and Science in Sport and Exercise*—the flagship journal of the American College of Sports Medicine—asks authors to justify sample sizes by reporting power calculations for the primary statistical tests, sample size was justified in only 35% of a random sample of studies from the journal (Twomey et al., 2021). However, merely justifying sample size using a power calculation is not always accurate or sufficiently transparent, for myriad reasons: there may be a mismatch between the statistical test identified in the power calculation (e.g., *t*-test) and the primary analysis performed in the study (e.g., group-by-time interaction from an ANOVA); the study might reference an inappropriate effect size (e.g., based on within-subject vs. between-subject comparisons); the study might rely on pilot data to estimate the population effect size; the study might improperly specify one-tail vs. two-tail tests; the study might assume a single outcome despite analyzing numerous dependent variables (i.e., failure to adjust alpha); the researchers might fail to account for anticipated participant attrition; the researchers might fail to account for testing of multiple dependent variables; and researchers might fail to report enough information to enable readers to replicate the calculations (Chan et al., 2008; Charles et al., 2009). To this latter point, a systematic review on the effects of sprint interval training found that 21 of 27 studies (78%) either did not report power calculations or failed to provide adequate information (Bonafiglia et al., 2022).

Poor statistical power is not a new problem. An analysis of statistical procedures, sample sizes, and significance levels of articles published in *Research Quarterly in Exercise and Sport* (volume 46, 1975) found that studies with small effects had little chance (<20%) of accurately rejecting the null hypothesis, with the actual statistical power ranging from 0.06 to 0.20 (Christensen & Christensen, 1977). In the contemporary exercise-related literature, small sample sizes and low statistical power are still pervasive. For instance, a recent meta-analysis on the physiological effects of high-intensity interval training (HIIT)—a paradigm which has become widely popular in both the science and practice of exercise—included 48 studies exhibiting a median sample size of just n=10 per group (Mattioni Maturana et al., 2021). Moreover, the small samples were used to assess what was found to be a small-to-medium effect (pooled effect size 0.40), resulting in most studies (88%) exhibiting statistical power in the range of 0–20%. This is similar in magnitude to around half of studies in biomedical sciences (Dumas-Mallet et al., 2017). Small-scale studies with poor statistical power are thought to result from the current research paradigm underpinned by perverse incentives (Higginson & Munafò, 2016).

The issue of low statistical power has generally not improved despite repeated examples of its deleterious consequences (Smaldino & McElreath, 2016).

Indeed, along with high sampling variability, the low statistical power often associated with small samples may explain the difficulty faced in study replication (Stanley et al., 2018). Aside from *a priori* power calculations, two additional factors should inform the selected sample size in kinesiology research. First, researchers must ‘oversample’ in anticipation of inevitable dropout/attrition, particularly in longitudinal training studies (Viken et al., 2019). Second, sample size calculations assume perfect measurements and do not account for the less-than-perfect reliability of most exercise measures (e.g., $\dot{V}O_2$max, blood pressure, self-reported measures of physical activity, other patient-reported outcomes). Measurement error entails substantial loss of statistical power that is rarely compensated for with larger samples (Baugh, 2002; Charter, 1997; Groenwold & Dekkers, 2020; Loken & Gelman, 2017). Recruiting larger samples, when deemed necessary for statistical robustness, can be difficult—especially in invasive and/or mechanistic studies or when funding and laboratory resources are limited. One possible solution is to incentivize collaboration among institutions to implement large-scale studies and to pool data collected at individual sites. While this initiative comes with many challenges (e.g., inconsistent laboratory personnel and equipment, subtle differences in data collection protocols), there is a clear benefit to statistical power, conferring greater confidence in the conclusions that are drawn.

4.7 Selective outcomes and cherry-picking. In exploratory studies, particularly those in which the concept of interest is inherently multidimensional (e.g., metabolism, immune function, executive function, health-related quality of life), it is common for researchers to measure numerous dependent variables, often across multiple time points. Even with statistics that correct for multiple comparisons, it is easy for authors to emphasize positive outcomes and overlook negative ones (Ioannidis, 2005). An analysis of sports nutrition/physiology research revealed that approximately 86% of studies failed to prioritize outcomes (Earnest et al., 2018), enabling greater flexibility for researchers to select the findings perceived to be the most novel, intriguing, or in line with expectations. By designating the primary outcome variables of interest in advance, the flexibility to differentially choose the ones that are most favorable can be attenuated. Failing to pre-register directional hypotheses and the appropriate statistical analyses gives researchers the opportunity to cherry-pick outcomes, perform expedient analyses (e.g., to experiment with various combinations of covariates), and present biased interpretations of results. Both hypothesis-generating (exploratory) and hypothesis testing (confirmatory) studies are integral components of kinesiology research (Bishop, 2008). Nevertheless, there is an increased need for authors to be explicit as to whether their studies are exploratory or confirmatory (Büttner et al., 2020; Caldwell et al., 2020).

Not accounting for placebo effects. Placebo-controlled trials are the benchmark of clinical research into new drugs, as well as being crucial for exploring the ergogenic effects of sports supplements and devices. However, failing to account for the inherent psychobiological effects of the placebo phenomenon itself can lead to an overestimation of real effects. For instance, placebo may contribute up to 25% of the total intervention effect of extracellular buffers and up to 59% of the total intervention effect of caffeine supplements (Marticorena et al., 2021). Placebo influences the subjective responses to pain (Colloca, 2019), the psychological effects of exercise training (Desharnais et al., 1993), and even the anticipated effects of altitude training interventions (Garvican et al., 2011). Accordingly, in addition to a placebo arm, interventional studies in kinesiology should employ a ‘no-intervention’ comparator group. Such a three-way study design will enable researchers to differentiate physiological and psychobiological effects (Marticorena et al., 2021).

5. Recommendations for Reform

Despite a growing body of work on the prevalence of research misconduct and QRPs in science, a systematic and quantitative exploration is needed to further elucidate the extent of the problem in kinesiology. Cross-discipline qualitative data capturing the experiences of researchers regarding QRPs would also be valuable. Reform must then follow a two-pronged approach that addresses the manifestations (symptoms) of QRPs in kinesiology, as well as the incentives (causes) that potentially give rise to them.

With respect to addressing the manifestations, the first step to greater transparency and accuracy in reporting hypotheses, methods, and results, is to expand study preregistration (Caldwell et al., 2020). By providing a framework with which to compare the registered trial and the published manuscript (Büttner et al., 2020), preregistration can help mitigate some of the methodological discrepancies between them (Chahal et al., 2012). Subsequently, registered reports generally outperform non-registered reports in methodological rigor, analytical rigor, and overall paper quality (Soderberg et al., 2021). Moreover, in clinical settings, mandatory registration leads to more transparent research and reliable data (Aslam et al., 2013). Preregistration of kinesiology research is not obligatory, and it therefore remains very rare (Twomey et al., 2021). As such, this should initially be incentivized (e.g., through the use of badges; Kidwell et al., 2016; Munafò et al., 2017) but may eventually need to be mandated. Academics have repeatedly called for authors of sports and exercise science-related research to register their hypotheses and methods prior to data collection (e.g., on publicly available repositories or by submitting registered reports) (Caldwell et al., 2020), and the *Journal of Sports Sciences* has announced its support for Open Science practices like study preregistration (Abt et al., 2022).

Although QRPs can still manifest in preregistered studies (e.g., in deviation from the registered protocol), a commitment to preregistration may help improve methodological rigor to a level approaching clinical research while also helping to improve the perceived legitimacy of kinesiology among other scientific fields. Embracing preregistration would also simplify the process of performing replication studies, thereby expediting the verification of results from exploratory research.

A simple and cost-effective means of eliciting the positive outcomes of preregistration without mandating it would be to require authors to simultaneously submit their manuscripts alongside the previously approved ethics or institutional review board applications. Reviewers or journal editorial assistants could then check for disparities between the aims and objectives stated in the two documents. Notwithstanding the additional administrative burden, such a system would improve transparency in study reporting and help mitigate the *post hoc* derivation or modification of study aims, particularly in interventional research.

As an adjunct to study preregistration, the enhanced use of preprint servers prior to formal submission for peer-review can benefit researchers, particularly those in the early stages of their careers, by affording rapid dissemination of study findings, increasing (open) accessibility, establishing priority or concurrence, and facilitating feedback from, and collaboration with, the academic community (Sarabipour et al., 2019). By guaranteeing the dissemination of methods and data among the scientific community, preprint servers may attenuate the publication pressures that underpin research misconduct and QRPs. Although the widespread use of preprints is associated with many challenges—and should be used cautiously in research relating to drugs, vaccines, or medical devices that directly affect the treatment of patients (Flanagin et al., 2020) —in most kinesiology research, the use of preprints may confer a net benefit.

As aforenoted, many QRPs manifest as inappropriate/incorrect statistical procedures. This issue can be addressed with robust reporting of inferential statistics that include, at the very least, effect sizes, confidence intervals, correction for familywise error rate, and designation of primary outcomes. More stringent levels of statistical significance may also be required. For instance, some scientific fields that depend on high levels of confidence (e.g., particle physics, genetics) have implemented significance levels of five-sigma to reduce false discovery rates (i.e., a pre-determined alpha of 3×10^{-7}). While these criteria are unnecessarily stringent for most kinesiology-related research, authors in our field have proposed more conservative pre-determined alphas of 0.01 or 0.001 instead of the standard 0.05 (Gandevia, 2021). The responsibility ultimately falls to journal editors and manuscript reviewers to adopt stronger policies and enforce more robust statistical reporting in the articles they accept for publication (Bernard, 2019).

In a further effort to overcome high false positivity rates, a disclosure-based solution for prospective manuscripts has been proposed, whereby authors and reviewers adhere to a checklist that ensures transparency in the description of methods and results, as well as in the manuscript review process (Simmons et al., 2011). When there are multiple outcomes in a study, Earnest *et al. (2018)* suggested that authors establish a prioritized analysis schema: to encourage authors to carefully prioritize the most important aspects of a study, to strengthen *a priori* analyses for future studies, and to help contextualize secondary or tertiary outcomes from exploratory testing. Establishing a hierarchy of aims should be determined during study conceptualization and design (Freemantle, 2001), and primary outcomes should be congruent with those considered in the power calculations (Andrade, 2015). The combined approach of more robust and/or conservative statistical reporting, and greater transparency in pre-determined aims and objectives, may eventually help rebuild trust in published research.

When considering the high risk of false positive results in research (Ioannidis, 2005), which is often congruent with weak experimental findings (Diong et al., 2018), it may be that impropriety arises more often through earnest error or statistical naiveté rather than through fraud or malintent (Steen, 2011), in accordance with the so-called "Hanlon’s razor." Indeed, self-reported research-misconduct scores were higher among younger researchers (postdoctoral fellows, assistant/associate professors) compared to their more experienced peers (Maggio et al., 2019). This may be due to heightened pressure for career advancement among younger scientists but perhaps also due to their relative ignorance of responsible research practices (Fanelli et al., 2015). In the biomedical and social sciences, funding bodies—such as the National Institutes of Health—mandate that all award recipients undergo research ethics training (DuBois et al., 2008). However, because most kinesiology studies are not externally funded, we must double our efforts to integrate good research and statistical practices into kinesiology-related higher-education programs. This may partially prevent the next generation of career scientists from committing avoidable QRPs.

Despite all of these cogent recommendations and perennial calls for improvements in science reporting, research misconduct and QRPs persist (Edwards & Roy, 2016). This suggests that treating the superficial manifestations through the aforementioned strategies may not be sufficient to evoke long-term behavior change. Reform may instead be achieved by addressing the incentives underlying misconduct and QRPs (i.e., the root causes). This will undoubtedly be a long and difficult process that requires large-scale buy-in—from academics, practitioners, publishers, and institutions—and long-term strategies that incrementally shift the emphasis away from quantitative performance metrics.

The drive to publish is immediate and self-perpetuating. This is particularly true for early-career scientists who are eager to accrue quantitative markers of achievement (Nosek et al., 2012) and secure promotion/tenure (Maggio et al., 2019). In fact, postdoctoral researchers and assistant professors report the highest scores for publication pressures, funding pressures, and competitiveness (Gopalakrishna et al., 2022). This can be partially addressed by urging researchers to pursue long-term projects that are both robust and that make a (relatively) profound contribution to science, rather than short-term, high-volume outputs of questionable integrity. The overarching aim should be to emphasize quality over quantity. It is essential to strike a balance in this regard because, according to Edwards and Roy (2016), an overemphasis on quality manifests as a stringent and overcautious system characterized by multiple blinded studies and mandatory replication of results. By contrast, a frivolous emphasis on quantity would sacrifice scientific rigor in both study design/execution and subsequent peer review, resulting in high error rates (Edwards & Roy, 2016). Both extremes would likely slow the attainment of new knowledge. Assuming the goal of research is to promote scientific progress, optimum productivity is likely to stem from carefully balancing research quantity and quality (Edwards & Roy, 2016) (**Figure 4**).

Knowledge dissemination and career progression currently rely almost exclusively on the (somewhat restrictive) framework provided by peer-reviewed academic publishing. While it is unrealistic to expect academics and their institutions to ever abandon the current publishing paradigm (such is the mutual dependency between journals and academia), meaningful change can be achieved by reforming the structure of, and the incentives underpinning, the current system. Specifically, “promoting truth over publishability” requires efforts to reduce the costs associated with sharing and accessing research; enhance the use of community-driven, open-access journals; and enhance the use of public repositories that enable continuous peer review (Nosek & Bar-Anan, 2012). This may enable a level of scrutiny that we hypothesize would yield better standards of practice in the discipline. It would also enhance our ability to meet the overarching purpose of public science: healthy knowledge accumulation. The proposed changes may also shift the emphasis away from quantitative metrics (e.g., impact factors) and toward knowledge-building incentives that yield research with a demonstrably meaningful impact on theory or practice. In other words, the central focus of research should be on altruistic values and the moral/ethical obligations of scientific enquiry to the society it serves. This strategy would need to be complemented with a system that supports the next generation of scientists in striking a healthy balance between internal accuracy motives (i.e., learning and publishing robust data) and purely professional ones (Nosek et al., 2012).

At the very least, moving the discipline toward these ambitious long-term objectives requires greater support for organizations and societies that are already striving to improve standards of reporting within kinesiology. For instance, the Society for Transparency, Openness, and Replication in Kinesiology (STORK) provides one of the few cooperative platforms for health and exercise scientists to improve their methods and practices. This is largely achieved by emphasizing research quality over quantity, striving for accuracy and transparency in the reporting of data and statistics, promoting alternative publishing models that ensure online accessibility over cost, and encouraging critical analysis of kinesiology research practices. It is vital that academics, practitioners, universities, research institutions, and journals devoted to kinesiology adopt initiatives like STORK and work collaboratively to embolden an ethos of transparency and openness in research. The reputation of kinesiology and related disciplines, and the integrity of the data produced therein, may depend on it.

Conclusions

There is an ongoing conflict between the desire to “advance as a respected scientist” and to retain one’s intellectual integrity in academia when surrounded by “perverse” incentives that legitimize and even mandate QRPs as a means of progression. Real progress, however, will be to acknowledge that these two ambitions need not be in opposition. This requires a fundamental change in what it means to “advance as a respected scientist” and how it is achieved. This paper outlines how the manifestations (symptoms) of QRPs can be addressed; namely by placing more emphasis on under- and postgraduate education relating to research design and the importance of robust statistical reporting, incentivizing and perhaps mandating study pre-registration in kinesiology research, and employing disclosure-based systems whereby authors and reviewers adhere to a checklist that ensures transparency in the description of methods and results. Nevertheless, the “perverse incentives” that typically underpin research misconduct and QRPs can only be addressed by nurturing a gradual change in the research paradigm—away from the current emphasis on quantitative performance metrics and toward a model that encourages transparency and openness and that fulfils altruistic values.

Declarations

Ethics approval and consent to participate: Not applicable

Consent for publication: Not applicable

Availability of data and material: Not applicable

Competing interests: None.

Funding: Not applicable

Authors' contributions: NBT conceived the manuscript idea. NBT and PE drafted and edited the manuscript. NBT and PE. approved the final version of the manuscript.

Acknowledgements: Not applicable

References

Abt, G., Boreham, C., Davison, G., Jackson, R., Nevill, A., Wallace, E., & Williams, M. (2020). Power, precision, and sample size estimation in sport and exercise science research. *Journal of Sports Sciences*, *38*(17), 1933–1935. https://doi.org/10.1080/02640414.2020.1776002

Abt, G., Jobson, S., Morin, J.-B., Passfield, L., Sampaio, J., Sunderland, C., & Twist, C. (2022). Raising the bar in sports performance research. *Journal of Sports Sciences*, *40*(2), 125–129. https://doi.org/10.1080/02640414.2021.2024334

Altman, D. G., Gore, S. M., Gardner, M. J., & Pocock, S. J. (1983). Statistical guidelines for contributors to medical journals. *British Medical Journal (Clinical Research Ed.)*, *286*(6376), 1489–1493.

American Psychological Association. (2020). *Publication manual of the American Psychological Association 2020: The official guide to APA style* (7th ed.).

Andersen, M. B., McCullagh, P., & Wilson, G. J. (2007). But what do the numbers really tell us? Arbitrary metrics and effect size reporting in sport psychology research. *Journal of Sport & Exercise Psychology*, *29*(5), 664–672. https://doi.org/10.1123/jsep.29.5.664

Andrade, C. (2015). The primary outcome measure and its importance in clinical trials. *The Journal of Clinical Psychiatry*, *76*(10), e1320-1323. https://doi.org/10.4088/JCP.15f10377

Aslam, A., Imanullah, S., Asim, M., & El-Menyar, A. (2013). Registration of clinical trials: Is it really needed? *North American Journal of Medical Sciences*, *5*(12), 713–715. https://doi.org/10.4103/1947-2714.123266

Atkinson, G., & Batterham, A. M. (2015). True and false interindividual differences in the physiological response to an intervention. *Experimental Physiology*, *100*(6), 577–588. https://doi.org/10.1113/EP085070

Atkinson, G., Loenneke, J. P., Fahs, C. A., Abe, T., & Rossow, L. M. (2015). Individual differences in the exercise-mediated blood pressure response: Regression to the mean in disguise? *Clinical Physiology and Functional Imaging*, *35*(6), 490–491. https://doi.org/10.1111/cpf.12211

Atkinson, G., & Nevill, A. M. (2001). Selected issues in the design and analysis of sport performance research. *Journal of Sports Sciences*, *19*(10), 811–827. https://doi.org/10.1080/026404101317015447

Atkinson, G., Williamson, P., & Batterham, A. M. (2019). Issues in the determination of “responders” and “non-responders” in physiological research. *Experimental Physiology*, *104*(8), 1215–1225. https://doi.org/10.1113/EP087712

Bachynski, K. E., & Smoliga, J. M. (2021). Pseudomedicine for sports concussions in the USA. *The Lancet. Neurology*, *20*(10), 791–792. https://doi.org/10.1016/S1474-4422(19)30250-9

Barlow, P., Serôdio, P., Ruskin, G., McKee, M., & Stuckler, D. (2018). Science organisations and Coca-Cola’s “war” with the public health community: Insights from an internal industry document. *Journal of Epidemiology and Community Health*, *72*(9), 761–763. https://doi.org/10.1136/jech-2017-210375

Baugh, F. (2002). Correcting effect sizes for score reliability: A reminder that measurement and substantive issues are linked inextricably. *Educational and*

Psychological Measurement, *62*(2), 254–263. https://doi.org/10.1177/0013164402062002004

Begley, C. G., & Ioannidis, J. P. A. (2015). Reproducibility in science: Improving the standard for basic and preclinical research. *Circulation Research*, *116*(1), 116–126. https://doi.org/10.1161/CIRCRESAHA.114.303819

Bernard, C. (2019). Changing the way we report, interpret, and discuss our results to rebuild trust in our research. *ENeuro*, *6*(4), ENEURO.0259-19.2019. https://doi.org/10.1523/ENEURO.0259-19.2019

Bernards, J. R., Sato, K., Haff, G. G., & Bazyler, C. D. (2017). Current research and statistical practices in sport science and a need for change. *Sports*, *5*(4), Article 4. https://doi.org/10.3390/sports5040087

Bes-Rastrollo, M., Schulze, M. B., Ruiz-Canela, M., & Martinez-Gonzalez, M. A. (2013). Financial conflicts of interest and reporting bias regarding the association between sugar-sweetened beverages and weight gain: A systematic review of systematic reviews. *PLoS Medicine*, *10*(12), e1001578; dicsussion e1001578. https://doi.org/10.1371/journal.pmed.1001578

Biagioli, M., & Lippman, A. (2020). *Gaming the metrics: Misconduct and manipulation in academic research*. MIT Press. https://doi.org/10.7551/mitpress/11087.001.0001

Bishop, D. (2008). An applied research model for the sport sciences. *Sports Medicine (Auckland, N.Z.)*, *38*(3), 253–263. https://doi.org/10.2165/00007256-200838030-00005

Bleakley, C., & MacAuley, D. (2002). The quality of research in sports journals. *British Journal of Sports Medicine*, *36*(2), 124–125. https://doi.org/10.1136/bjsm.36.2.124

Blumenthal, J. A., Smith, P. J., Mabe, S., Hinderliter, A., Lin, P.-H., Liao, L., Welsh-Bohmer, K. A., Browndyke, J. N., Kraus, W. E., Doraiswamy, P. M., Burke, J. R., & Sherwood, A. (2019). Lifestyle and neurocognition in older adults with cognitive impairments: A randomized trial. *Neurology*, *92*(3), e212–e223. https://doi.org/10.1212/WNL.0000000000006784

Bonafiglia, J. T., Islam, H., Preobrazenski, N., & Gurd, B. J. (2022). Risk of bias and reporting practices in studies comparing VO2max responses to sprint interval vs. continuous training: A systematic review and meta-analysis. *Journal of Sport and Health Science*, *11*(5), 552–566. https://doi.org/10.1016/j.jshs.2021.03.005

Brainard, J. (2020). Articles in “predatory” journals receive few or no citations. *Science (New York, N.Y.)*, *367*(6474), 129. https://doi.org/10.1126/science.367.6474.129

Brischoux, F., & Angelier, F. (2015). Academia’s never-ending selection for productivity. *Scientometrics*, *103*(1), 333–336. https://doi.org/10.1007/s11192-015-1534-5

Büttner, F., Toomey, E., McClean, S., Roe, M., & Delahunt, E. (2020). Are questionable research practices facilitating new discoveries in sport and exercise medicine? The proportion of supported hypotheses is implausibly high. *British Journal of Sports Medicine*, *54*(22), 1365–1371. https://doi.org/10.1136/bjsports-2019-101863

Caldwell, A. R., Vigotsky, A. D., Tenan, M. S., Radel, R., Mellor, D. T., Kreutzer, A., Lahart, I. M., Mills, J. P., Boisgontier, M. P., & Consortium for Transparency in Exercise Science (COTES) Collaborators. (2020). Moving Sport and Exercise Science Forward: A Call for the Adoption of More Transparent Research

Practices. *Sports Medicine (Auckland, N.Z.)*, *50*(3), 449–459. https://doi.org/10.1007/s40279-019-01227-1

Callaway, E. (2011). Fraud investigation rocks Danish university. *Nature*. https://doi.org/10.1038/news.2011.703

Chahal, J., Tomescu, S. S., Ravi, B., Bach, B. R., Ogilvie-Harris, D., Mohamed, N. N., & Gandhi, R. (2012). Publication of sports medicine-related randomized controlled trials registered in ClinicalTrials.gov. *The American Journal of Sports Medicine*, *40*(9), 1970–1977. https://doi.org/10.1177/0363546512448363

Chan, A.-W., Hróbjartsson, A., Jørgensen, K. J., Gøtzsche, P. C., & Altman, D. G. (2008). Discrepancies in sample size calculations and data analyses reported in randomised trials: Comparison of publications with protocols. *BMJ (Clinical Research Ed.)*, *337*, a2299. https://doi.org/10.1136/bmj.a2299

Charles, P., Giraudeau, B., Dechartres, A., Baron, G., & Ravaud, P. (2009). Reporting of sample size calculation in randomised controlled trials: Review. *BMJ (Clinical Research Ed.)*, *338*, b1732. https://doi.org/10.1136/bmj.b1732

Charter, R. A. (1997). Effect on measurement error on tests of statistical significance. *Journal of Clinical and Experimental Neuropsychology*, *19*(3), 458–462. https://doi.org/10.1080/01688639708403872

Christensen, J. E., & Christensen, C. E. (1977). Statistical power analysis of health, physical education, and recreation research. *Research Quarterly*, *48*(1), 204–208. https://doi.org/10.1080/10671315.1977.10762173

Colloca, L. (2019). The Placebo Effect in Pain Therapies. *Annual Review of Pharmacology and Toxicology*, *59*, 191–211. https://doi.org/10.1146/annurev-pharmtox-010818-021542

Coursol, A., & Wagner, E. E. (1986). Effect of positive findings on submission and acceptance rates: A note on meta-analysis bias. *Professional Psychology: Research and Practice*, *17*(2), 136–137. https://doi.org/10.1037/0735-7028.17.2.136

Cumming, G. (2008). Replication and p Intervals: P Values Predict the Future Only Vaguely, but Confidence Intervals Do Much Better. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *3*(4), 286–300. https://doi.org/10.1111/j.1745-6924.2008.00079.x

Cumming, G., Fidler, F., Kalinowski, P., & Lai, J. (2012). The statistical recommendations of the American psychological association publication manual: Effect sizes, confidence intervals, and meta-analysis. *Australian Journal of Psychology*, *64*, 138–146. https://doi.org/10.1111/j.1742-9536.2011.00037.x

Curran-Everett, D. (2016). Explorations in statistics: Statistical facets of reproducibility. *Advances in Physiology Education*, *40*(2), 248–252. https://doi.org/10.1152/advan.00042.2016

Curran-Everett, D., & Benos, D. J. (2007). Guidelines for reporting statistics in journals published by the American Physiological Society: The sequel. *Advances in Physiology Education*, *31*(4), 295–298. https://doi.org/10.1152/advan.00022.2007

Dahlberg, J. E., & Mahler, C. C. (2006). The poehlman case: Running away from the truth. *Science and Engineering Ethics*, *12*(1), 157–173. https://doi.org/10.1007/s11948-006-0016-9

Dalton, R. (2005). Obesity expert owns up to million-dollar crime. *Nature*, *434*(7032), Article 7032. https://doi.org/10.1038/434424a

Del Coso, J., Lara, B., Ruiz-Moreno, C., & Salinero, J. J. (2019). Challenging the Myth of Non-Response to the Ergogenic Effects of Caffeine Ingestion on Exercise Performance. *Nutrients*, *11*(4), E732. https://doi.org/10.3390/nu11040732

Desharnais, R., Jobin, J., Côté, C., Lévesque, L., & Godin, G. (1993). Aerobic exercise and the placebo effect: A controlled study. *Psychosomatic Medicine*, *55*(2), 149–154. https://doi.org/10.1097/00006842-199303000-00003

Diels, J., Cunha, M., Manaia, C., Sabugosa-Madeira, B., & Silva, M. (2011). Association of financial or professional conflict of interest to research outcomes on health risks or nutritional assessment studies of genetically modified products. *Food Policy*, *36*(2), 197–203. https://doi.org/10.1016/j.foodpol.2010.11.016

Dinis-Oliveira, R. J., & Magalhães, T. (2016). The Inherent Drawbacks of the Pressure to Publish in Health Sciences: Good or Bad Science. *F1000Research*, *4*, 419. https://doi.org/10.12688/f1000research.6809.2

Diong, J., Butler, A. A., Gandevia, S. C., & Héroux, M. E. (2018). Poor statistical reporting, inadequate data presentation and spin persist despite editorial advice. *PloS One*, *13*(8), e0202121. https://doi.org/10.1371/journal.pone.0202121

DuBois, J. M., Dueker, J. M., Anderson, E. E., & Campbell, J. (2008). The development and assessment of an NIH-funded research ethics training program. *Academic Medicine: Journal of the Association of American Medical Colleges*, *83*(6), 596–603. https://doi.org/10.1097/ACM.0b013e3181723095

Dumas-Mallet, E., Button, K. S., Boraud, T., Gonon, F., & Munafò, M. R. (2017). Low statistical power in biomedical science: A review of three human research domains. *Royal Society Open Science*, *4*(2), 160254. https://doi.org/10.1098/rsos.160254

Earnest, C. P., Roberts, B. M., Harnish, C. R., Kutz, J. L., Cholewa, J. M., & Johannsen, N. M. (2018). Reporting Characteristics in Sports Nutrition. *Sports*, *6*(4), Article 4. https://doi.org/10.3390/sports6040139

Edwards, M. A., & Roy, S. (2016). Academic Research in the 21st Century: Maintaining Scientific Integrity in a Climate of Perverse Incentives and Hypercompetition. *Environmental Engineering Science*, *34*(1), 51–61. https://doi.org/10.1089/ees.2016.0223

Eisenmann, J. (2017). Translational Gap between Laboratory and Playing Field: New Era to Solve Old Problems in Sports Science. *Translational Journal of the American College of Sports Medicine*, *2*(8), 37–43. https://doi.org/10.1249/TJX.0000000000000032

Ekkekakis, P., & Tiller, N. B. (2022). Extraordinary Claims in the Literature on High-Intensity Interval Training: II. Are the Extraordinary Claims Supported by Extraordinary Evidence? *Kinesiology Review*, *1*(aop), 1–14. https://doi.org/10.1123/kr.2022-0003

Errington, T. M., Mathur, M., Soderberg, C. K., Denis, A., Perfito, N., Iorns, E., & Nosek, B. A. (2021). Investigating the replicability of preclinical cancer biology. *ELife*, *10*, e71601. https://doi.org/10.7554/eLife.71601

Fanelli, D. (2009). How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PloS One*, *4*(5), e5738. https://doi.org/10.1371/journal.pone.0005738

Fanelli, D. (2010). “Positive” Results Increase Down the Hierarchy of the Sciences. *PLOS ONE*, *5*(4), e10068. https://doi.org/10.1371/journal.pone.0010068

Fanelli, D., Costas, R., & Larivière, V. (2015). Misconduct Policies, Academic Culture and Career Stage, Not Gender or Pressures to Publish, Affect Scientific Integrity. *PloS One*, *10*(6), e0127556. https://doi.org/10.1371/journal.pone.0127556

Fisher, R. A. (1926). The arrangement of field experiments. *Journal of the Ministry of Agriculture*, *33*, 503–515. https://doi.org/10.23637/rothamsted.8v61q

Flanagin, A., Fontanarosa, P. B., & Bauchner, H. (2020). Preprints Involving Medical Research-Do the Benefits Outweigh the Challenges? *JAMA*, *324*(18), 1840–1843. https://doi.org/10.1001/jama.2020.20674

Franco, A., Malhotra, N., & Simonovits, G. (2014). Publication bias in the social sciences: Unlocking the file drawer. *Science*, *345*(6203), 1502–1505. https://doi.org/10.1126/science.1255484

Freemantle, N. (2001). Interpreting the results of secondary end points and subgroup analyses in clinical trials: Should we lock the crazy aunt in the attic? *BMJ*, *322*(7292), 989–991. https://doi.org/10.1136/bmj.322.7292.989

Gandevia, S. (2021). Publications, replication and statistics in physiology plus two neglected curves. *The Journal of Physiology*, *599*(6), 1719–1721. https://doi.org/10.1113/JP281360

Gardner, M. J., Altman, D. G., Jones, D. R., & Machin, D. (1983). Is the statistical assessment of papers submitted to the “British Medical Journal” effective? *Br Med J (Clin Res Ed)*, *286*(6376), 1485–1488. https://doi.org/10.1136/bmj.286.6376.1485

Garvican, L. A., Pottgiesser, T., Martin, D. T., Schumacher, Y. O., Barras, M., & Gore, C. J. (2011). The contribution of haemoglobin mass to increases in cycling performance induced by simulated LHTL. *European Journal of Applied Physiology*, *111*(6), 1089–1101. https://doi.org/10.1007/s00421-010-1732-z

Gaspar, D. E. P., & Esteves, M. D. L. (2021). Awareness of the Misconduct in Sports Science Research. *Annals of Applied Sport Science*, *9*(3), 0–0. https://doi.org/10.52547/aassjournal.934

Gasparyan, A. Y., Ayvazyan, L., Akazhanov, N. A., & Kitas, G. D. (2014). Self-correction in biomedical publications and the scientific impact. *Croatian Medical Journal*, *55*(1), 61–72. https://doi.org/10.3325/cmj.2014.55.61

Gerrits, R. G., Jansen, T., Mulyanto, J., Berg, M. J. van den, Klazinga, N. S., & Kringos, D. S. (2019). Occurrence and nature of questionable research practices in the reporting of messages and conclusions in international scientific Health Services Research publications: A structured assessment of publications authored by researchers in the Netherlands. *BMJ Open*, *9*(5), e027903. https://doi.org/10.1136/bmjopen-2018-027903

Gibbs, K. D., & Griffin, K. A. (2013). What Do I Want to Be with My PhD? The Roles of Personal Values and Structural Dynamics in Shaping the Career Interests of Recent Biomedical Science PhD Graduates. *CBE Life Sciences Education*, *12*(4), 711–723. https://doi.org/10.1187/cbe.13-02-0021

Gibbs, N. M., & Gibbs, S. V. (2015). Misuse of “trend” to describe “almost significant” differences in anaesthesia research. *British Journal of Anaesthesia*, *115*(3), 337–339. https://doi.org/10.1093/bja/aev149

Goodman, S. N. (1992). A comment on replication, P-values and evidence. *Statistics in Medicine*, *11*(7), 875–879. https://doi.org/10.1002/sim.4780110705

Gopalakrishna, G., Ter Riet, G., Vink, G., Stoop, I., Wicherts, J. M., & Bouter, L. M. (2022). Prevalence of questionable research practices, research misconduct and their potential explanatory factors: A survey among academic researchers in The Netherlands. *PloS One*, *17*(2), e0263023. https://doi.org/10.1371/journal.pone.0263023

Groenwold, R. H. H., & Dekkers, O. M. (2020). Measurement error in clinical research, yes it matters. *European Journal of Endocrinology*, *183*(3), E3–E5. https://doi.org/10.1530/EJE-20-0550

Halperin, I., Vigotsky, A. D., Foster, C., & Pyne, D. B. (2018). Strengthening the Practice of Exercise and Sport-Science Research. *International Journal of Sports Physiology and Performance*, *13*(2), 127–134. https://doi.org/10.1123/ijspp.2017-0322

Harris, J. D., Cvetanovich, G., Erickson, B. J., Abrams, G. D., Chahal, J., Gupta, A. K., McCormick, F. M., & Bach, B. R. (2014). Current status of evidence-based sports medicine. *Arthroscopy: The Journal of Arthroscopic & Related Surgery: Official Publication of the Arthroscopy Association of North America and the International Arthroscopy Association*, *30*(3), 362–371. https://doi.org/10.1016/j.arthro.2013.11.015

Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., & Rafols, I. (2015). Bibliometrics: The Leiden Manifesto for research metrics. *Nature*, *520*(7548), 429–431. https://doi.org/10.1038/520429a

Higginson, A. D., & Munafò, M. R. (2016). Current Incentives for Scientists Lead to Underpowered Studies with Erroneous Conclusions. *PLOS Biology*, *14*(11), e2000995. https://doi.org/10.1371/journal.pbio.2000995

Hopkins, W. G., Marshall, S. W., Batterham, A. M., & Hanin, J. (2009). Progressive Statistics for Studies in Sports Medicine and Exercise Science. *Medicine &*

Science in Sports & Exercise, *41*(1), 3–12. https://doi.org/10.1249/MSS.0b013e31818cb278

How to spot dodgy academic journals. (2020). *The Economist*. https://www.economist.com/graphic-detail/2020/05/30/how-to-spot-dodgy-academic-journals

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine*, *2*(8), e124. https://doi.org/10.1371/journal.pmed.0020124

Ioannidis, J. P. A., Boyack, K. W., & Baas, J. (2020). Updated science-wide author databases of standardized citation indicators. *PLoS Biology*, *18*(10), e3000918. https://doi.org/10.1371/journal.pbio.3000918

Ioannidis, J. P. A., Ntzani, E. E., Trikalinos, T. A., & Contopoulos-Ioannidis, D. G. (2001). Replication validity of genetic association studies. *Nature Genetics*, *29*(3), Article 3. https://doi.org/10.1038/ng749

Islam, H., & Gurd, B. J. (2020). Exercise response variability: Random error or true differences in exercise response? *Experimental Physiology*, *105*(12), 2022–2024. https://doi.org/10.1113/EP089015

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the Prevalence of Questionable Research Practices With Incentives for Truth Telling. *Psychological Science*, *23*(5), 524–532. https://doi.org/10.1177/0956797611430953

Jones, S. R., Carley, S., & Harrison, M. (2003). An introduction to power and sample size estimation. *Emergency Medicine Journal*, *20*(5), 453–458. https://doi.org/10.1136/emj.20.5.453

Kardeş, S., Levack, W., Özkuk, K., Atmaca Aydın, E., & Seringeç Karabulut, S. (2020). Retractions in Rehabilitation and Sport Sciences Journals: A Systematic

Review. *Archives of Physical Medicine and Rehabilitation*, *101*(11), 1980–1990. https://doi.org/10.1016/j.apmr.2020.03.010

Kerr, N. L. (1998). HARKing: Hypothesizing After the Results are Known. *Personality and Social Psychology Review*, *2*(3), 196–217. https://doi.org/10.1207/s15327957pspr0203_4

Khatra, O., Shadgan, A., Taunton, J., Pakravan, A., & Shadgan, B. (2021). A bibliometric analysis of the top cited articles in sports and exercise medicine. *Orthopaedic Journal of Sports Medicine*, *9*(1), 2325967120969902. https://doi.org/10.1177/2325967120969902

Kidwell, M. C., Lazarević, L. B., Baranski, E., Hardwicke, T. E., Piechowski, S., Falkenberg, L.-S., Kennett, C., Slowik, A., Sonnleitner, C., Hess-Holden, C., Errington, T. M., Fiedler, S., & Nosek, B. A. (2016). Badges to Acknowledge Open Practices: A Simple, Low-Cost, Effective Method for Increasing Transparency. *PLoS Biology*, *14*(5), e1002456. https://doi.org/10.1371/journal.pbio.1002456

King, E. G., Oransky, I., Sachs, T. E., Farber, A., Flynn, D. B., Abritis, A., Kalish, J. A., & Siracuse, J. J. (2018). Analysis of retracted articles in the surgical literature. *American Journal of Surgery*, *216*(5), 851–855. https://doi.org/10.1016/j.amjsurg.2017.11.033

Kintisch, E. (2006). *Poehlman Sentenced to 1 Year of Prison*. Science.Org. https://www.science.org/content/article/poehlman-sentenced-1-year-prison

Kiss, A., Temesi, Á., Tompa, O., Lakner, Z., & Soós, S. (2021). Structure and trends of international sport nutrition research between 2000 and 2018: Bibliometric mapping of sport nutrition science. *Journal of the International Society of Sports Nutrition*, *18*(1), 12. https://doi.org/10.1186/s12970-021-00409-5

Klimentidis, Y. C., Raichlen, D. A., Bea, J., Garcia, D. O., Wineinger, N. E., Mandarino, L. J., Alexander, G. E., Chen, Z., & Going, S. B. (2018). Genome-wide association study of habitual physical activity in over 377,000 UK Biobank participants identifies multiple variants including CADM2 and APOE. *International Journal of Obesity (2005)*, *42*(6), 1161–1176. https://doi.org/10.1038/s41366-018-0120-3

Knudson, D. (2009). Significant and meaningful effects in sports biomechanics research. *Sports Biomechanics*, *8*(1), 96–104. https://doi.org/10.1080/14763140802629966

Knudson, D. (2012). Twenty-year trends of authorship and sampling in applied biomechanics research. *Perceptual and Motor Skills*, *114*(1), 16–20. https://doi.org/10.2466/11.PMS.114.1.16-20

Knudson, D. (2017a). Twenty Years of Authorship, Sampling, and References in Kinesiology Research Reports. *International Journal of Kinesiology in Higher Education*, *1*(2), 44–52. https://doi.org/10.1080/24711616.2017.1282760

Knudson, D. (2017b). Confidence crisis of results in biomechanics research. *Sports Biomechanics*, *16*(4), 425–433. https://doi.org/10.1080/14763141.2016.1246603

Knudson, D. (2019). Judicious Use of Bibliometrics to Supplement Peer Evaluations of Research in Kinesiology in: Kinesiology Review Volume 8 Issue 2 (). *Kinesiology Review*, *8*(2), 100–109. https://doi.org/doi.org/10.1123/kr.2017-0046

Kondro, W. (2005). Université de Montréal in the dark about fraud. *CMAJ*, *172*(10), 1278-1278-a. https://doi.org/10.1503/cmaj.050455

Krans, B. (2022). *Lawsuit: Coca-Cola Fake Ads About Obesity*. Healthline.Com. https://www.healthline.com/health/coca-cola-false-advertising-unhealthy-drinks

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*, 863. https://doi.org/10.3389/fpsyg.2013.00863

Landhuis, E. (2016). Scientific literature: Information overload. *Nature*, *535*(7612), Article 7612. https://doi.org/10.1038/nj7612-457a

Lesser, L. I., Ebbeling, C. B., Goozner, M., Wypij, D., & Ludwig, D. S. (2007). Relationship between funding source and conclusion among nutrition-related scientific articles. *PLoS Medicine*, *4*(1), e5. https://doi.org/10.1371/journal.pmed.0040005

Lohse, K. R., Sainani, K. L., Taylor, J. A., Butson, M. L., Knight, E. J., & Vickers, A. J. (2020). Systematic review of the use of “magnitude-based inference” in sports science and medicine. *PLOS ONE*, *15*(6), e0235318. https://doi.org/10.1371/journal.pone.0235318

Loken, E., & Gelman, A. (2017). Measurement error and the replication crisis. *Science (New York, N.Y.)*, *355*(6325), 584–585. https://doi.org/10.1126/science.aal3618

Lucas, M. (2015). Conflicts of interest in nutritional sciences: The forgotten bias in meta-analysis. *World Journal of Methodology*, *5*(4), 175–178. https://doi.org/10.5662/wjm.v5.i4.175

Maggio, L., Dong, T., Driessen, E., & Artino, A. (2019). Factors associated with scientific misconduct and questionable research practices in health professions education. *Perspectives on Medical Education*, *8*(2), 74–82. https://doi.org/10.1007/s40037-019-0501-x

Mallapaty, S. (2020). China bans cash rewards for publishing papers. *Nature*, *579*(7797), 18. https://doi.org/10.1038/d41586-020-00574-8

Marticorena, F. M., Carvalho, A., Oliveira, L. F. D., Dolan, E., Gualano, B., Swinton, P., & Saunders, B. (2021). Nonplacebo Controls to Determine the Magnitude of Ergogenic Interventions: A Systematic Review and Meta-analysis. *Medicine and Science in Sports and Exercise*, *53*(8), 1766–1777. https://doi.org/10.1249/MSS.0000000000002635

Martin, J. J. (2020). Grants: The Good, the Bad, the Ugly, and the Puzzling. *Kinesiology Review*, *10*(1), 18–28. https://doi.org/10.1123/kr.2020-0013

Mattioni Maturana, F., Martus, P., Zipfel, S., & NIEß, A. M. (2021). Effectiveness of HIIE versus MICT in Improving Cardiometabolic Risk Factors in Health and Disease: A Meta-analysis. *Medicine and Science in Sports and Exercise*, *53*(3), 559–573. https://doi.org/10.1249/MSS.0000000000002506

Montero, D., & Lundby, C. (2017). Refuting the myth of non-response to exercise training: 'Non-responders' do respond to higher dose of training. *The Journal of Physiology*, *595*(11), 3377–3387. https://doi.org/10.1113/JP273480

Moylan, E. C., & Kowalczuk, M. K. (2016). Why articles are retracted: A retrospective cross-sectional study of retraction notices at BioMed Central. *BMJ Open*, *6*(11), e012047. https://doi.org/10.1136/bmjopen-2016-012047

Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., du Sert, N. P., Simonsohn, U., Wagenmakers, E.-J., Ware, J. J., & Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, *1*, 0021. https://doi.org/10.1038/s41562-016-0021

Murphy, J., Mesquida, C., Caldwell, A. R., Earp, B. D., & Warne, J. P. (2023). Proposal of a Selection Protocol for Replication of Studies in Sports and Exercise Science. *Sports Medicine (Auckland, N.Z.)*. https://doi.org/10.1007/s40279-022-01749-1

Nagele, P. (2003). Misuse of standard error of the mean (sem) when reporting variability of a sample. A critical evaluation of four anaesthesia journals. *BJA: British Journal of Anaesthesia*, *90*(4), 514–516. https://doi.org/10.1093/bja/aeg087

Nosek, B. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716. https://doi.org/10.1126/science.aac4716

Nosek, B. A., & Bar-Anan, Y. (2012). Scientific Utopia: I. Opening scientific communication. *ArXiv:1205.1055 [Physics]*. http://arxiv.org/abs/1205.1055

Nosek, B. A., Spies, J. R., & Motyl, M. (2012). Scientific Utopia: II. Restructuring Incentives and Practices to Promote Truth Over Publishability. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *7*(6), 615–631. https://doi.org/10.1177/1745691612459058

Open Science Collaboration. (2015). PSYCHOLOGY. Estimating the reproducibility of psychological science. *Science (New York, N.Y.)*, *349*(6251), aac4716. https://doi.org/10.1126/science.aac4716

Pabst, S., Brand, M., & Wolf, O. T. (2013). Stress effects on framed decisions: There are differences for gains and losses. *Frontiers in Behavioral Neuroscience*, *7*, 142. https://doi.org/10.3389/fnbeh.2013.00142

Padilla, J., Leary, E., & Limberg, J. K. (2021). Identifying responders versus non-responders: Incorporation of controls is required for sound statistical inference. *Experimental Physiology*, *106*(2), 375–376. https://doi.org/10.1113/EP089142

Prom-Wormley, E., Adkins, A., Waldman, I. D., & Dick, D. (2017). Critical Issues in Genetic Association Studies. In *Psychological Science Under Scrutiny* (pp. 221–249). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119095910.ch12

Quan, W., Chen, B., & Shu, F. (2017). Publish or impoverish: An investigation of the monetary reward system of science in China (1999-2016). *Aslib Journal of Information Management*, *69*(5), 486–502. https://doi.org/10.1108/AJIM-01-2017-0014

Robertson, P. S. (2020). Man & machine: Adaptive tools for the contemporary performance analyst. *Journal of Sports Sciences*, *38*(18), 2118–2126. https://doi.org/10.1080/02640414.2020.1774143

Sabroe, I., Dockrell, D. H., Vogel, S. N., Renshaw, S. A., Whyte, M. K. B., & Dower, S. K. (2007). Identifying and hurdling obstacles to translational research. *Nature Reviews. Immunology*, *7*(1), 77–82. https://doi.org/10.1038/nri1999

Sainani, K., & Chamari, K. (2022). Wish List for Improving the Quality of Statistics in Sport Science. *International Journal of Sports Physiology and Performance*, *17*(5), 673–674. https://doi.org/10.1123/ijspp.2022-0023

Sainani, K. L., Borg, D. N., Caldwell, A. R., Butson, M. L., Tenan, M. S., Vickers, A. J., Vigotsky, A. D., Warmenhoven, J., Nguyen, R., Lohse, K. R., Knight, E. J., & Bargary, N. (2021). Call to increase statistical collaboration in sports science, sport and exercise medicine and sports physiotherapy. *British Journal of Sports Medicine*, *55*(2), 118–122. https://doi.org/10.1136/bjsports-2020-102607

Sarabipour, S., Debat, H. J., Emmott, E., Burgess, S. J., Schwessinger, B., & Hensel, Z. (2019). On the value of preprints: An early career researcher perspective. *PLoS Biology*, *17*(2), e3000151. https://doi.org/10.1371/journal.pbio.3000151

Scheel, A. M., Schijen, M., & Lakens, D. (2020). *An excess of positive results: Comparing the standard Psychology literature with Registered Reports*. PsyArXiv. https://doi.org/10.31234/osf.io/p6e9c

Schmid, S. L. (2017). Five years post-DORA: Promoting best practices for research assessment. *Molecular Biology of the Cell*, *28*(22), 2941–2944. https://doi.org/10.1091/mbc.E17-08-0534

Schulz, K. F., Altman, D. G., Moher, D., & CONSORT Group. (2010). CONSORT 2010 Statement: Updated guidelines for reporting parallel group randomised trials. *BMC Medicine*, *8*, 18. https://doi.org/10.1186/1741-7015-8-18

Schulz, R., Langen, G., Prill, R., Cassel, M., & Weissgerber, T. L. (2022). Reporting and transparent research practices in sports medicine and orthopaedic clinical trials: A meta-research study. *BMJ Open*, *12*(8), e059347. https://doi.org/10.1136/bmjopen-2021-059347

Serodio, P., Ruskin, G., McKee, M., & Stuckler, D. (2020). Evaluating Coca-Cola’s attempts to influence public health ‘in their own words’: Analysis of Coca-Cola emails with public health academics leading the Global Energy Balance Network. *Public Health Nutrition*, *23*, 1–7. https://doi.org/10.1017/S1368980020002098

Serra-Garcia, M., & Gneezy, U. (2021). Nonreplicable publications are cited more than replicable ones. *Science Advances*, *7*(21), eabd1705. https://doi.org/10.1126/sciadv.abd1705

Shen, C., & Björk, B.-C. (2015). “Predatory” open access: A longitudinal study of article volumes and market characteristics. *BMC Medicine*, *13*, 230. https://doi.org/10.1186/s12916-015-0469-2

Silberzahn, R., Uhlmann, E. L., Martin, D. P., Anselmi, P., Aust, F., Awtrey, E., Bahník, Š., Bai, F., Bannard, C., Bonnier, E., Carlsson, R., Cheung, F., Christensen, G., Clay, R., Craig, M. A., Dalla Rosa, A., Dam, L., Evans, M. H., Flores Cervantes, I., … Nosek, B. A. (2018). Many Analysts, One Data Set: Making Transparent How Variations in Analytic Choices Affect Results. *Advances in Methods and Practices in Psychological Science*, *1*(3), 337–356. https://doi.org/10.1177/2515245917747646

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. *Psychological Science*, *22*(11), 1359–1366. https://doi.org/10.1177/0956797611417632

Smaldino, P. E., & McElreath, R. (2016). The Natural Selection of Bad Science. *Royal Society Open Science*, *3*(9), 160384. https://doi.org/10.1098/rsos.160384

Smoliga, J. M. (2020). Interpreting Biomarker Data After Concussion and Repeated Subconcussive Head Impacts: Challenges in Evaluating Brain Protection. *JAMA Neurology*, *77*(12), 1477–1478. https://doi.org/10.1001/jamaneurol.2020.3467

Soderberg, C. K., Errington, T. M., Schiavone, S. R., Bottesini, J., Thorn, F. S., Vazire, S., Esterling, K. M., & Nosek, B. A. (2021). Initial evidence of research quality of registered reports compared with the standard publishing model. *Nature Human Behaviour*, *5*(8), Article 8. https://doi.org/10.1038/s41562-021-01142-4

Sox, H. C., & Rennie, D. (2006). Research Misconduct, Retraction, and Cleansing the Medical Literature: Lessons from the Poehlman Case. *Annals of Internal Medicine*, *144*(8), 609–613. https://doi.org/10.7326/0003-4819-144-8-200604180-00123

Spence, J. C., & Blanchard, C. (2001). Publication bias in sport and exercise psychology: The games we play. *International Journal of Sport Psychology*, *32*(4), 386–399.

Stanley, T. D., Carter, E. C., & Doucouliagos, H. (2018). What meta-analyses reveal about the replicability of psychological research. *Psychological Bulletin*, *144*(12), 1325–1346. https://doi.org/10.1037/bul0000169

Statista. (2020). *Global sports nutrition & supplement market 2025*. Statista. https://www.statista.com/statistics/450168/global-sports-nutrition-market/

Steen, R. G. (2011). Misinformation in the medical literature: What role do error and fraud play? *Journal of Medical Ethics*, *37*(8), 498–503. https://doi.org/10.1136/jme.2010.041830

Sterling, T. D., Rosenbaum, W. L., & Weinkam, J. J. (1995). Publication Decisions Revisited: The Effect of the Outcome of Statistical Tests on the Decision to Publish and Vice Versa. *The American Statistician*, *49*(1), 108–112. https://doi.org/10.2307/2684823

Stuckler, D., Ruskin, G., & McKee, M. (2018). Complexity and conflicts of interest statements: A case-study of emails exchanged between Coca-Cola and the principal investigators of the International Study of Childhood Obesity, Lifestyle and the Environment (ISCOLE). *Journal of Public Health Policy*, *39*(1), 49–56. https://doi.org/10.1057/s41271-017-0095-7

Teo, Y. Y. (2008). Common statistical issues in genome-wide association studies: A review on power, data quality control, genotype calling and population structure. *Current Opinion in Lipidology*, *19*(2), 133–143. https://doi.org/10.1097/MOL.0b013e3282f5dd77

The Society for Transparency, Openness, and Replication in Kinesiology. (n.d.). *Reports in Sport and Exercise*. Retrieved March 3, 2022, from https://storkjournals.org/index.php/rrik

Thomas, J. R., Salazar, W., & Landers, D. M. (1991). What is missing in p less than .05? Effect size. *Research Quarterly for Exercise and Sport*, *62*(3), 344–348. https://doi.org/10.1080/02701367.1991.10608733

Thompson, B. (1996). AERA editorial policies regarding statistical significance testing: Three suggested reforms. *Educational Researcher*, *25*(2), 26–30. https://doi.org/10.2307/1176337

Tijdink, J. K., Verbeke, R., & Smulders, Y. M. (2014). Publication Pressure and Scientific Misconduct in Medical Scientists. *Journal of Empirical Research on Human Research Ethics*, *9*(5), 64–71. https://doi.org/10.1177/1556264614552421

Tiller, N. B. (2020). *The Skeptic's Guide to Sports Science: Confronting Myths of the Health and Fitness Industry*. Routledge.

Tiller, N. B., Sullivan, J. P., & Ekkekakis, P. (2022). Baseless Claims and Pseudoscience in Health and Wellness: A Call to Action for the Sports, Exercise, and Nutrition-Science Community. *Sports Medicine (Auckland, N.Z.)*. https://doi.org/10.1007/s40279-022-01702-2

Twomey, R., Yingling, V., Warne, J., Schneider, C., McCrum, C., Atkins, W., Murphy, J., Medina, C. R., Harlley, S., & Caldwell, A. (2021). The Nature of Our

Literature: A Registered Report on the Positive Result Rate and Reporting Practices in Kinesiology. *Communications in Kinesiology*, *1*(3), Article 3. https://doi.org/10.51224/cik.v1i3.43

Vagenas, G., Palaiothodorou, D., & Knudson, D. (2018). Thirty-year Trends of Study Design and Statistics in Applied Sports and Exercise Biomechanics Research. *International Journal of Exercise Science*, *11*(1), 239–259.

van de Schoot, R., Winter, S. D., Griffioen, E., Grimmelikhuijsen, S., Arts, I., Veen, D., Grandfield, E. M., & Tummers, L. G. (2021). The Use of Questionable Research Practices to Survive in Academia Examined With Expert Elicitation, Prior-Data Conflicts, Bayes Factors for Replication Effects, and the Bayes Truth Serum. *Frontiers in Psychology*, *12*, 621547. https://doi.org/10.3389/fpsyg.2021.621547

Van Noorden, R. (2010). Metrics: A profusion of measures. *Nature*, *465*(7300), 864–866. https://doi.org/10.1038/465864a

Van Noorden, R., & Singh Chawla, D. (2019). Hundreds of extreme self-citing scientists revealed in new database. *Nature*, *572*(7771), 578–579. https://doi.org/10.1038/d41586-019-02479-7

Vankov, I., Bowers, J., & Munafò, M. R. (2014). Article Commentary: On the Persistence of Low Power in Psychological Science. *Quarterly Journal of Experimental Psychology*, *67*(5), 1037–1040. https://doi.org/10.1080/17470218.2014.885986

Viken, H., Reitlo, L. S., Zisko, N., Nauman, J., Aspvik, N. P., Ingebrigtsen, J. E., Wisløff, U., & Stensvold, D. (2019). Predictors of Dropout in Exercise Trials in Older Adults: The Generation 100 Study. *Medicine & Science in Sports & Exercise*, *51*(1), 49–55. https://doi.org/10.1249/MSS.0000000000001742

Walters, J. (2015, August 11). Nutrition experts alarmed by nonprofit downplaying role of junk food in obesity. *The Guardian*. https://www.theguardian.com/society/2015/aug/11/obesity-junk-food-exercise-global-energy-balance-network-coca-cola

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA Statement on p-Values: Context, Process, and Purpose. *The American Statistician*, *70*(2), 129–133. https://doi.org/10.1080/00031305.2016.1154108

Watanabe, R. M. (2011). Statistical issues in gene association studies. *Methods in Molecular Biology (Clifton, N.J.)*, *700*, 17–36. https://doi.org/10.1007/978-1-61737-954-3_2

Wiedermann, C. J. (2016). Ethical publishing in intensive care medicine: A narrative review. *World Journal of Critical Care Medicine*, *5*(3), 171–179. https://doi.org/10.5492/wjccm.v5.i3.171

Williams, C. J., Li, Z., Harvey, N., Lea, R. A., Gurd, B. J., Bonafiglia, J. T., Papadimitriou, I., Jacques, M., Croci, I., Stensvold, D., Wisloff, U., Taylor, J. L., Gajanand, T., Cox, E. R., Ramos, J. S., Fassett, R. G., Little, J. P., Francois, M. E., Hearon, C. M., … Coombes, J. S. (2021). Genome wide association study of response to interval and continuous exercise training: The Predict-HIIT study. *Journal of Biomedical Science*, *28*(1), 37. https://doi.org/10.1186/s12929-021-00733-7

Wood, J., Freemantle, N., King, M., & Nazareth, I. (2014). Trap of trends to statistical significance: Likelihood of near significant P value becoming more significant with extra data. *BMJ*, *348*, g2215. https://doi.org/10.1136/bmj.g2215

Yong, E. (2012). Nobel laureate challenges psychologists to clean up their act. *Nature*. https://doi.org/10.1038/nature.2012.11535

Figures and Tables

Figure. 1. Scopus® data showing the chronological trend in the number of articles published in “Sports Science” journals (all subject areas and all regions/countries) from 2000 to 2021, the number of articles published in the top 50 highest-ranked “Sports Science” journals (listed in ascending order of the average number of weighted citations for a given year relative to the documents [*n*] published in the journal in the three previous years), and the total number of “Sports Science” journals.

Figure 2. Self-citation rates of 554 of the top-cited “Sport Sciences” researchers, worldwide. The mean ± SD percentage of self-citations was 14 ± 7% and the median (IQR) was 13% (8%) (range 0% - 54%). Six authors exhibited self-citation rates that exceeded three standard deviations of the mean: 54%, 44%, 42%, 41%, 37%, and 35%, respectively. In the database as a whole (all disciplines, $n = 195,605$ records), the mean self-citation rate was 13 ± 9% and the median (IQR) was 12% (10%). Data acquired from public records **(Ioannidis et al., 2020).**

Figure 3. The probability that a replicated experiment will achieve $p < 0.05$ and the 80% prediction intervals for the *p* value given by a replication. The line graph depicts the probability that a repeated experiment will successfully replicate the initial *p*-value (data from Goodman, 1992 [(Goodman, 1992)]; Curran-Everett, 2016 [(Curran-Everett, 2016)]). For example, an initial experiment obtaining a *p*-value of 0.001 would have ~90% chance of being replicated with a $p < 0.05$. However, with an initial *p*-vlaue of 0.05, the probability of a successful replication falls to ~50%.

The grey bars and shading represent the intervals that include the *p*-value given by a replication with an 80% chance (secondary *y*-axis) (data are from Cumming, 2008 [(Cumming, 2008)]). For example, if an initial experiment obtains $p = 0.05$, the 80% prediction interval for replication in a duplicate experiment will range from 0.00008 to 0.44. Redrawn from Gandevia (Gandevia, 2021).

Figure 4. Assuming the overarching goal of research is to maximize scientific progress, optimum productivity is likely to be found by carefully balancing research quantity and quality. Too much emphasis on quality will lead to a loss of productivity, while too much emphasis on quantity will lead to increased error rate. *Productivity loss owing to human error, misconduct, etc. Redrawn from Edwards and Roy (2016).

Table 1. Characteristics of sports medicine studies, including the percentage of studies reporting supported hypotheses. Reproduced from Büttner et al. (2020).

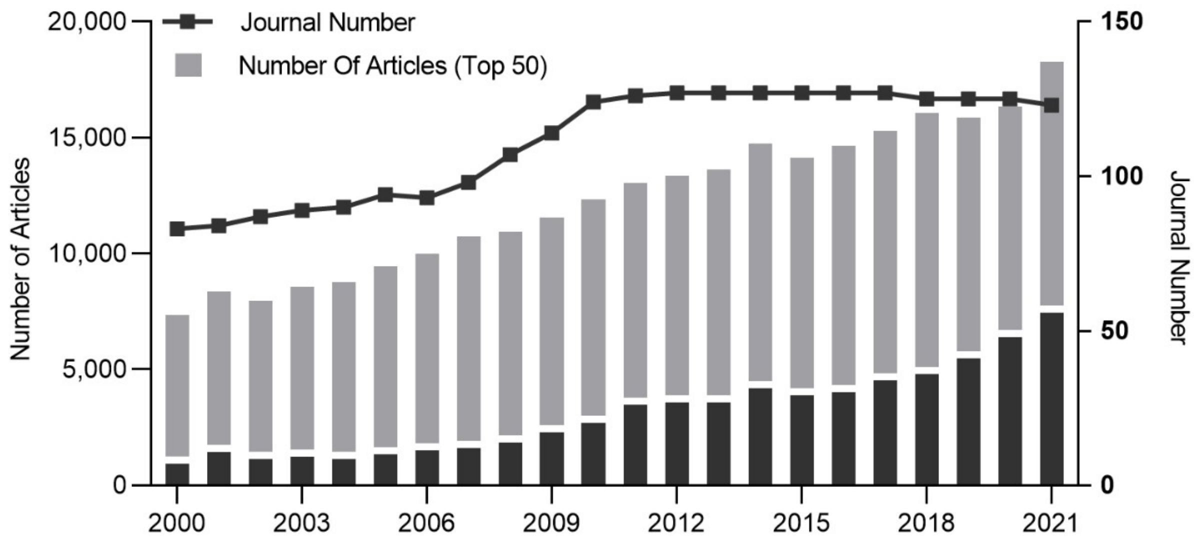
Journal Number
Number Of Articles (Top 50)
20,000
15,000
10,000
5,000
0
Number of Articles
150
100
50
0
Journal Number
2000
2003
2006
2009
2012
2015
2018
2021

Fig. 1.

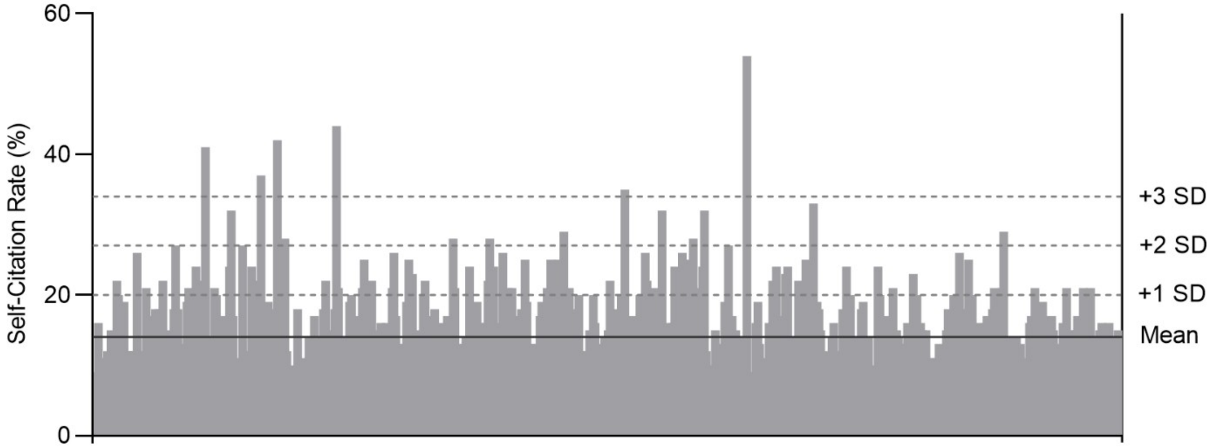

Fig. 2.

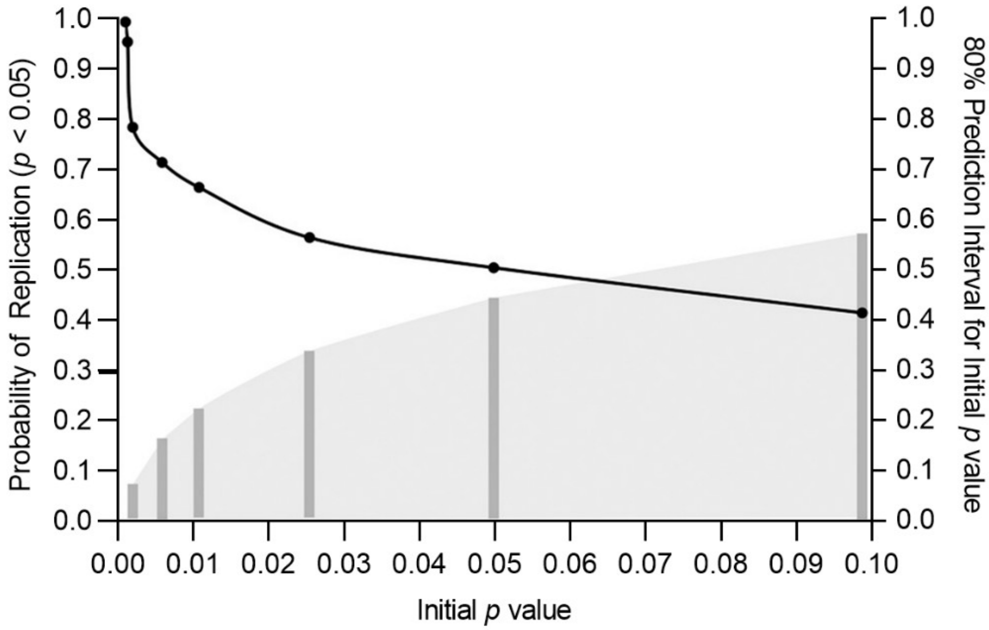
Probability of Replication (p < 0.05)
80% Prediction Interval for Initial p value
Initial p value
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0.00 0.01 0.02 0.03 0.04 0.05 0.06 0.07 0.08 0.09 0.10

Fig. 3.

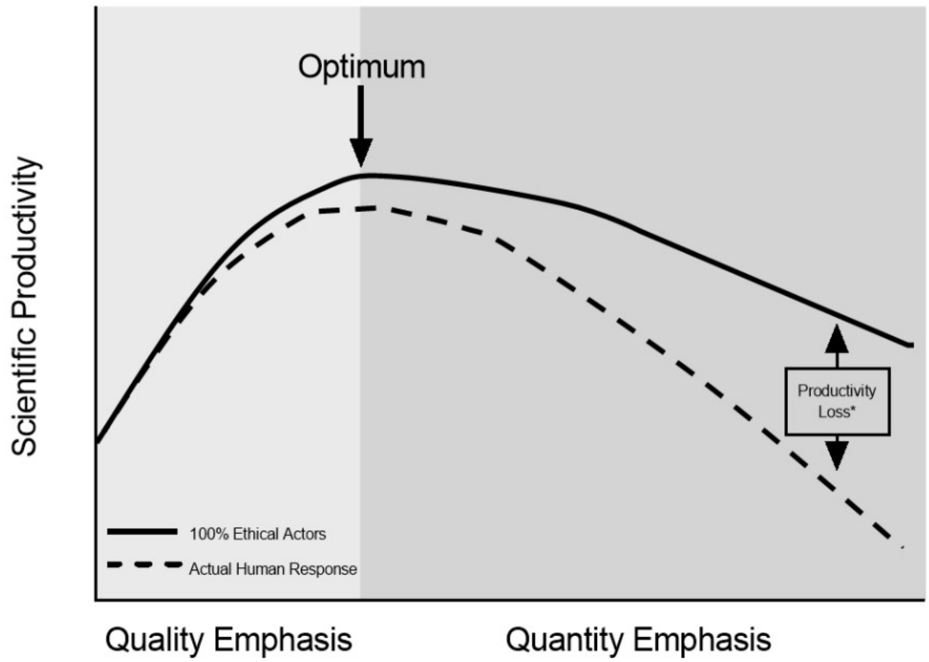
Optimum
Scientific Productivity
Productivity Loss*
100% Ethical Actors
Actual Human Response
Quality Emphasis
Quantity Emphasis

Fig. 4.

Table 1. Characteristics of sports medicine studies, including the percentage of studies reporting supported hypotheses. Reproduced from Büttner *et al*. (Büttner et al., 2020).

	n	**%**	**95% CI**
Original research	215	-	
Not reporting a hypothesis	86	40	33.5 - 46.5
Reporting at least one hypothesis	129	60	53.5 - 66.5
Original research reporting at least one hypothesis	129	-	
Stating an alternative primary hypothesis	102	79.1	72.0 – 86.1
Stating a null primary hypothesis	19	14.7	8.6 – 20.8
Stating both bull and alternative hypotheses within the	8	6.2	2.0 – 10.4
Original research reporting an alternative primary	102	-	
That state the direction of the hypothesis	91	89.2	83.2 – 95.2
That do not state the direction of the hypothesis	11	10.8	4.8 – 16.8
Is the primary hypothesis accepted or rejected?	129	-	
‘Accepted’ hypothesis	106	82.2	75.6 – 88.8
‘Rejected’ hypothesis	23	17.8	11.2 – 24.4
Is the accepted primary hypothesis fully or partially	106	-	
Fully	75	70.8	62.1 – 79.4
Partially	28	26.4	18.0 – 34.8
Unclear	3	2.8	0.1 – 6.0

n = number of studies