

Are questionable research practices facilitating new discoveries in sport and exercise medicine? The proportion of supported hypotheses is implausibly high

Fionn Büttner ¹, Elaine Toomey,² Shane McClean,¹ Mark Roe ¹, Eamonn Delahunt ^{1,3}

ABSTRACT

Questionable research practices (QRPs) are intentional and unintentional practices that can occur when designing, conducting, analysing, and reporting research, producing biased study results. Sport and exercise medicine (SEM) research is vulnerable to the same QRPs that pervade the biomedical and psychological sciences, producing false-positive results and inflated effect sizes. Approximately 90% of biomedical research reports supported study hypotheses, provoking suspicion about the field-wide presence of systematic biases to facilitate study findings that confirm researchers' expectations. In this education review, we introduce three common QRPs (ie, *HARKing*, *P-hacking* and *Cherry-picking*), perform a cross-sectional study to assess the proportion of original SEM research that reports supported study hypotheses, and draw attention to existing solutions and resources to overcome QRPs that manifest in exploratory research. We hypothesised that $\geq 85\%$ of original SEM research studies would report supported study hypotheses. Two independent assessors systematically identified, screened, included, and extracted study data from original research articles published between 1 January 2019 and 31 May 2019 in the *British Journal of Sports Medicine*, *Sports Medicine*, the *American Journal of Sports Medicine*, and the *Journal of Orthopaedic & Sports Physical Therapy*. We extracted data relating to whether studies reported that the primary hypothesis was supported or rejected by the results. Study hypotheses, methodologies, and analysis plans were preregistered at the Open Science

Framework. One hundred and twenty-nine original research studies reported at least one study hypothesis, of which 106 (82.2%) reported hypotheses that were supported by study results. Of 106 studies reporting that primary hypotheses were supported by study results, 75 (70.8%) studies reported that the primary hypothesis was fully supported by study results. The primary study hypothesis was partially supported by study results in 28 (26.4%) studies. We detail open science practices and resources that aim to safe-guard against QRPs that bely the credibility and replicability of original research findings.

PREAMBLE

Sport and exercise medicine (SEM), including sports physiotherapy, is a young research field with enormous scope for novel discoveries.¹⁻³ Most published SEM research is exploratory research - that is, research that does not transparently specify study aims, hypotheses, methodologies, and statistical analysis plans prior to data collection.^{2 4-7} Exploratory research, compared with confirmatory research that adheres to and reports prespecified study intentions, aims to generate new discoveries that advance clinical science and practice. Both exploratory and confirmatory research serve important purposes in the innovation and corroboration of SEM knowledge. However, problems can arise when exploratory research is falsely reported as confirmatory,^{4 8 9} increasing the probability that research findings are inaccurate, or worse still, false.¹⁰

Questionable research practices (QRPs) are intentional and unintentional practices that can occur when designing, conducting, analysing, and reporting research, producing biased study results.¹¹ QRPs are a frequent by-product of exploratory research that is falsely presented as confirmatory and increase the likelihood that study findings will be novel but also misleading.^{10 12} One-third of

scientists admit to using QRPs such as *P-hacking*, selective outcome reporting, and hypothesising after the results are known (*HARKing*) to generate statistically significant results.¹³ QRPs can occur independently or coexist in a research study.^{10 12 14 15}

SEM research is vulnerable to the same QRPs that pervade the biomedical and psychological sciences, producing false-positive results and inflated study effect sizes.^{10 14 16} Of all orthopaedic interventions comparing surgery to a non-operative alternative, only 20% are supported by at least one randomised controlled trial at low risk of bias.¹⁷ Of published SEM trials that are preregistered, approximately 35% exhibit discrepancies (eg, changes to statistical analyses and alteration or non-reporting of trial outcomes) between the preregistered protocol and the published manuscript.¹⁸ There is a scarcity of empirical meta-research that investigates the potential presence and burden of QRPs in SEM research.^{2 3 6 7 17-19} In this education review, we introduce three common QRPs¹¹ (ie, *P-hacking*, *Cherry-picking* and *HARKing*) and perform a cross-sectional study to assess the proportion of published SEM research studies that report supported hypotheses. Finally, we draw attention to existing solutions and resources to overcome QRPs that manifest in exploratory research.

Questionable or appropriate? An example of QRPs

A research team complete data collection for a prospective study that investigates the recovery of balance impairments in acutely concussed collision sport athletes. All athletes were tested in the preseason and those who sustained a sport-related concussion during the season were repeatedly tested post-injury. Upon data analysis, concussed athletes' post-injury balance scores are not statistically or clinically different compared with their preseason balance scores. The researchers perform dozens of regression analyses using different independent and outcome variables to assess whether preseason balance scores are associated with any musculoskeletal injury throughout the competitive season. In each regression analysis, the researchers trial numerous combinations of independent variables relating to balance, confounding variables such as sex, age, and concussion history, and outcome variables pertaining to any possible injury. One of many analyses discovers that athletes with poorer preseason balance scores are at greater risk of sustaining a

¹School of Public Health, Physiotherapy and Sports Science, University College Dublin—National University of Ireland, Dublin, Ireland

²School of Allied Health, University of Limerick, Limerick, Ireland

³Institute for Sport and Health, University College Dublin, Dublin, Ireland

Correspondence to Mr Fionn Büttner, School of Public Health, Physiotherapy and Sports Science, University College Dublin—National University of Ireland, Dublin, Dublin 4, Ireland; fionn.cleirigh-buttner@ucdconnect.ie

sport-related concussion than athletes with better preseason balance scores. This study has just transitioned from a recovery study to a prediction study. The researchers omit non-significant analyses and outcome measures and report only the significant analysis including relevant independent, confounding, and outcome variables. The researchers hypothesise that preseason balance impairments are associated with sport-related concussion, selectively cite relevant literature to support their findings, and plausibly explain their supported hypothesis. The paper is published in a reputable sports medicine journal.

P-hacking through the garden of forking paths (without leaving a trace)

Researchers face endless decisions when processing and analysing data that are collected during a research study.^{10 14 20} These decisions are diverse and range from whether to include or exclude outlying data points, to decisions about how many, and which, variables to include in a statistical analysis.^{12 15 21} Very often, decisions are made using arbitrary criteria (eg, based on ‘gut feeling’) or are determined after perusing the results.^{11 22} It is common for researchers to explore multiple analyses with only subtle differences between them when answering the same research question.¹⁴ For example, analyses may apply different combinations of independent, confounding, and outcome variables using different statistical tests with varying test parameters.¹⁴ By doing so, researchers engage in *P*-hacking, exploiting a ‘garden of forking paths’ whereby numerous data processing and analysis approaches are attempted, interpreted, and then finalised only when the most novel, desirable and often statistically significant ($p < 0.05$) result is obtained (figure 1).^{14 20 21}

In the above example, the researchers trial many regression analyses with different combinations of independent, confounding, and outcome variables. The researchers then use their knowledge of the result of each analysis to inform the parameters of their subsequent analysis and the eventual result they report.^{10 12} This undisclosed flexibility in exploratory data analysis can lead to a wide distribution of different results using the same data.^{23 24} For example, a regression analysis containing 20 independent binary variables will produce >1 million different results by including every level of each variable in all possible combinations (online supplementary file).²⁵ This multifold approach to statistical analysis

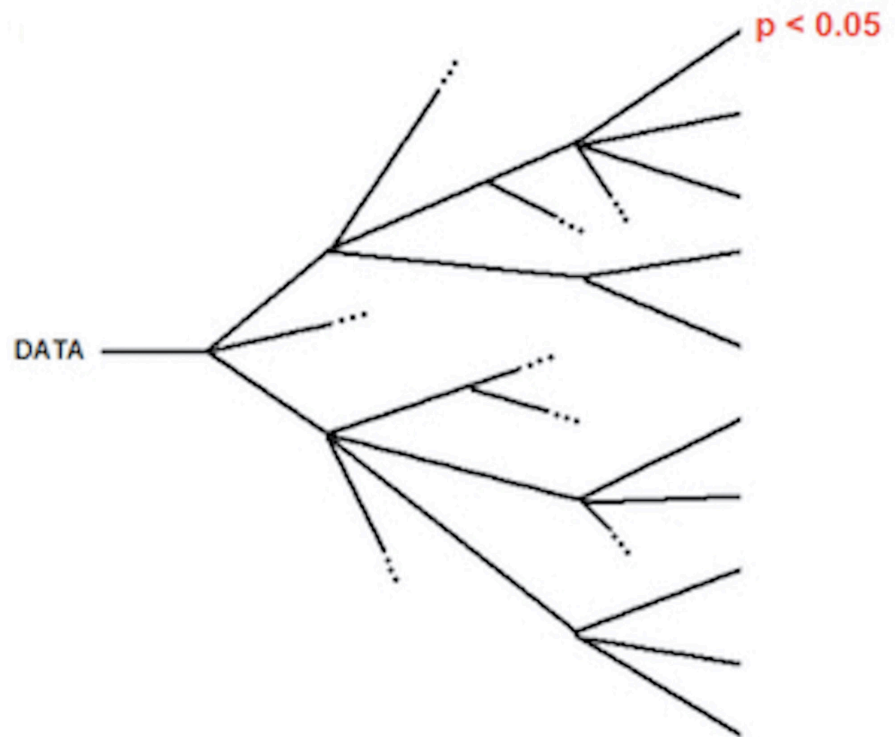


Figure 1 The ‘garden of forking paths’, whereby data are flexibly reprocessed and reanalysed using different strategies to obtain a statistically significant result (reproduced with permission: https://www.discovermagazine.com/mind/the-f-problem-with-the-p-value-sciences#.XX-e_50ZLNQI).

provides researchers with many options from which to select their most desirable study result. Trialling many data processing and analysis approaches and selecting and reporting only the method that produces the most impressive result ($p < 0.05$) creates a bias that will invariably favour inflated and potentially false research findings.^{12 15}

Cherry-picking in the orchard of statistically significant outcomes

Researchers frequently aim to ‘collect data for everything’ when designing a research study and when initiating data collection. Researchers may justify collecting data for as many independent variables and outcome variables as possible by claiming that they pursue groundbreaking findings in exploratory research areas.^{5 14} However, outcome sets quickly compound and become very large when researchers investigate many outcomes using numerous outcome measures across multiple study time-points with different outcome metrics.²⁶

Due to the retrospective nature of scientific reporting, researchers can selectively include and exclude study outcomes that demonstrate desirable or undesirable conclusions after they have observed the

results for each outcome.^{11 27} Selective outcome reporting, or outcome switching or ‘cherry-picking’ as it is also known, refers to the practice of using multiple outcomes in a research study but reporting only a selection. Selective outcome reporting increases the probability that a statistically significant study finding is due to chance.^{10 21 28} For example, the probability that one outcome variable will demonstrate a statistically significant result ($p < 0.05$) by chance (when the null hypothesis is true) is 5%. However, the probability that 1 outcome variable out of 13 tested outcome variables will achieve a statistically significant result, by chance (when the null hypothesis is true), is 49% (online supplementary file).^{10 29 30} When non-significant outcome variables are not reported in the study manuscript, it is impossible for readers to know to adjust their interpretation of significant findings.²⁹

In the above example, the researchers selectively report the significant association between poorer pre-season balance scores and subsequent risk of sustaining a sport-related concussion. The researchers fail to report the non-significant association between pre-season balance scores and subsequent (1) hip and groin injury

risk, (2) hamstring injury risk, (3) knee injury risk, and (4) ankle injury risk. The researchers test multiple outcomes but report only one, which negatively affects the probability that the reported significant association is true.^{10 28 31 32}

HARK the herald angels sing, glory to the exploratory finding

'Hypothesising After the Results are Known' (*HARKing*) is the behaviour of generating a study hypothesis to explain the result(s) of a research study—whether statistically significant or non-significant—only after the data have been analysed and the results are known.^{33–35} To *HARK*, researchers test multiple hypotheses after the data have been collected and when they identify the most desirable result, they hypothesise how such findings materialised as though they were expected in advance.³³ The study hypothesis is then falsely presented as though it was generated prior to data collection. SEM research is not immune to the threat of *HARKing*, which may be as high as 50% in other fields.^{36 37}

In the above example, the research team tests multiple hypotheses; from assessing the recovery of balance following sport-related concussion, to determining the association between pre-season balance impairments and in-season sport-related concussion (or any injury for that matter). The researchers eventually identify their most novel result—poorer preseason balance scores in collision sport athletes are associated with sport-related concussion—and report that they anticipated this finding before initiating the research study. The researchers *HARK* by explaining how they expected this result prior to data collection despite not actually expecting it at all; falsely presenting this exploratory finding as confirmatory.¹⁵

Every dataset contains statistically significant findings that occur solely by chance and do not represent true phenomena.^{4 29} Endlessly testing different hypotheses using the same data will eventually produce significant findings but at the increased risk of being a false-positive result.³⁵ Researchers can combine their knowledge of a content area with selectively chosen literature to retrospectively propose seemingly plausible hypotheses that can explain any research finding, however implausible they may actually be. Despite the perceived novelty and importance of exploratory research findings, biological implausibility, low prestudy odds, and the use of *HARKing* mean that many highly-exciting research findings

may occur due to chance and are unlikely to replicate.¹⁰

Is scientist behaviour simply a product of the academic environment: playing the game because they cannot change the rules?

While it can be easy to attribute QRPs to intentional, nefarious behaviour on behalf of researchers,^{38 39} less deceitful explanations are more common. Intricate decisions in the design, conduct, and analysis of research are frequently equivocal and not guided by definitive rules. As a result, decisions about the design of research or the analysis of data can be unclear and are often experience-, knowledge- and resource-dependent.²⁰

Academic culture prioritises novel research and positive, rather than negative, study findings.⁴⁰ Consequently, researchers are more likely to design studies and develop hypotheses that strive to confirm, rather than falsify, a theory.⁴⁰ Hindsight bias refers to the belief that an outcome (such as a surprising study finding) is inevitable after the outcome occurs.⁴¹ Researchers (and readers) can be vulnerable to hindsight bias by assuming that a previously unconsidered hypothesis, or a desirable/novel study finding, was inevitable only after the data were analysed, irrespective of how much flexibility or exploration was required to generate it.^{14 42}

Due to publication bias, novel and significant research findings are more likely to appear in the published literature than less exciting, non-significant findings.^{43 44} There is a high cost to pay for researchers who produce non-significant results due to a '*publish positively or perish!*' mindset in academic science.⁴⁵ In response, researchers must prioritise their most compelling study findings to increase their probability of publication.⁴⁶ When researchers observe non-significant or uninteresting results, they are motivated by career incentives to seek new relationships in the data, identify significant results, and report novel findings.^{38 47 48} Many researchers will consider QRPs an organic and integral part of the hypothesis-generation phase of research, without realising its detrimental effects on the validity of subsequent research findings.¹²

Near-flawless accuracy in the prediction of study findings: should scientists become gamblers?

Approximately 90% of biomedical research reports supported study

hypotheses—that is, hypotheses that are supported by study results. Such high predictive accuracy provokes suspicion about the field-wide presence of questionable research practices (QRPs) to facilitate study findings that confirm researchers' expectations. We performed a cross-sectional investigation to estimate the proportion of published, original sport and exercise medicine (SEM) research studies that report supported hypotheses. We hypothesised that $\geq 85\%$ of published, original research studies in SEM would report supported hypotheses. We also aimed to determine the proportion of published, original SEM research studies that report at least one study hypothesis. We hypothesised that approximately 50% of published, original research studies in SEM would report at least one study hypothesis.

METHODS

We searched PubMed to identify all content published between 1 January 2019 and 31 May 2019 in three of the highest impact factor SEM journals as per the 2018 Thomson Reuters Journal Citation Reports: *British Journal of Sports Medicine (BJSM)*, *Sports Medicine and the American Journal of Sports Medicine (AJSM)*, as well as one of the highest impact factor sports physiotherapy/physical therapy journals; *Journal of Orthopedic & Sports Physical Therapy (JOSPT)*. Two independent assessors screened title and abstracts, and full-text articles where necessary, to include only original research studies. We excluded study designs and publication types that did not fulfil the criteria of an original research study (eg, systematic review±meta-analysis, narrative review, education review, consensus statement, editorial, commentary, research update). We also excluded case-reports, qualitative studies, animal studies, cadaveric studies, and cellular/histological studies.

We assessed: (1) the proportion of original research studies reporting at least one study hypothesis and (2) the proportion of original research studies reporting a supported hypothesis. If studies reported at least one study hypothesis, we extracted meta-data relating to the following:

1. whether the primary study hypothesis was an alternative or null hypothesis;
2. whether the alternative primary study hypothesis was directional or not;
3. whether the study reported that the primary hypothesis was 'supported' or rejected by the results, and;

Table 1 Study characteristics including the proportion of studies reporting supported hypotheses

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

k, number of studies.

4. whether the supported study hypothesis was fully or partially supported by the study results.

Additional study information and materials (including the preregistered study protocol, study search strategy, data extraction methodology, definitions for extracted variables and study data) are available on the Open Science Framework (<https://osf.io/u43yc/>).

RESULTS

We identified 669 research items that were published in *BJSM*, *Sports Medicine*, *AJSM*, and *JOSPT* during our specified timeline. We included 215 (32.1%) eligible original research studies. Across 215 included original research studies, 129 (60%) reported at least one study hypothesis (online supplementary box 1). Of 129 original research studies reporting at least one study hypothesis, 106 (82.2%) studies reported a primary hypothesis that was supported by study results. Of 106 studies reporting that primary hypotheses were supported by study results, 75 (70.8%) studies reported that the primary hypothesis was fully supported by study results. The primary hypothesis was partially supported by study results in 28 (26.4%) studies. Full study results are included in [table 1](#) (including count and proportion data with 95% Confidence Intervals) and online supplementary tables 1–4.

DISCUSSION

We found that only 60% (k=129) of original SEM research studies published in *BJSM*, *Sports Medicine*, *AJSM*, and *JOSPT*

reported a study hypothesis. Popperian philosophy proposes that researchers should use study hypotheses to predict the study result they justifiably expect and then try to falsify this hypothesis through empirical investigation.⁴⁹ Only through failed refutation can a study hypothesis and its associated theory demonstrate robustness against falsifiability. The majority of SEM research may be exploratory, highlighted by the finding that only 60% (k=129) of published, original research studies reported a study hypothesis. Authors should aim to report and differentiate exploratory and confirmatory hypotheses with associated rationale, or lack thereof, for each. Without a stated study hypothesis, it is impossible to know whether researchers had any prestudy expectations or whether they simply explored the data to find whatever they could.

When studies reported a hypothesis, approximately 82% reported hypotheses that were either fully (71%) or partially (26%) supported by study results. Although the results of this study refute our own study hypothesis, this estimate (82%) is similar to the proportion of supported hypotheses reported in original research studies published in clinical biomedicine (89%) and psychology (91%–97%).^{40 50–52} The high proportion of supported hypotheses that we identified in the current study could be due to the excellent content knowledge that researchers possess and the genuine ability of researchers to correctly predict eventual study results. However, it is more likely

that setting broad and vague hypotheses (that easily garner evidential support), and using QRPs—whether intentionally or unintentionally—to accumulate evidence in favour of researchers’ prestudy beliefs, facilitate results that support study hypotheses.⁴⁰

Our investigation is limited by our selection of a convenience sample of original research studies published in only high impact factor SEM and sport physiotherapy journals (range: 3.058–11.645) from January 2019 until June 2019 (online supplementary file 1). The methodological quality of original research studies published in high impact factor SEM journals and the perceived expectations of authors about what type of study findings are more publishable in, and appeal to readers of, high impact factor SEM journals may introduce a selection bias that misrepresents SEM research.^{6 53–55}

Solutions to protect researchers against...themselves!

Transparently, rather than selectively, reporting planned hypotheses, outcome measures, and statistical analyses enables the scientific community to infer the extent of exploration that was undertaken to generate research findings.^{4 29} Preregistering study intentions provides a reference standard against which to compare the published manuscript and assess for deviations from what was prespecified. When preregistration is not performed, researchers should transparently report each study hypothesis, method, and analysis in the published manuscript and

Box 1 Open science solutions and resources to overcome questionable research practices

Preregistration

Preregistration is the practice of making a publicly accessible, time-stamped record of a research plan prior to data collection.⁵⁸ A preregistered study protocol should be specific, precise, and exhaustive by detailing study intentions including primary and secondary research questions and associated study hypotheses, methodologies, and statistical analyses.^{15 29} A publicly accessible, preregistered study protocol provides a clear distinction between confirmatory and exploratory research and allows readers to identify deviations in the final published manuscript from study intentions that preceded data collection. Researchers are still free to undertake exploratory research, but an available preregistered study protocol allows readers to differentiate study procedures that were intended before data collection and study procedures that were undertaken following data collection.

Preregistration databases such as clinicaltrials.gov, protocols.io, and the Open Science Framework (osf.io) offer platforms that store preregistered study protocols. Because preregistrations uploaded to platforms such as the Open Science Framework are time-stamped, preregistered study protocols can remain private and hidden from public view for a predetermined time-period and made publicly available at a desired time-point (e.g., upon manuscript submission to a journal).

Registered Reports

Researchers are often concerned that they will be unable to publish preregistered studies that do not identify novel and significant results. New initiatives exist that incentivise study preregistration with a view towards mandating publication.^{29 59 60} Using a Registered Report, researchers describe intended study research questions, hypotheses, methodologies, and statistical analysis plans prior to data collection, similar to conventional preregistration.⁶¹ Researchers submit their Registered Report to a journal that accommodates this format prior to study initiation and data collection. Following a stage of peer review that precedes data collection, an “accept” or “reject” decision is granted prior to data collection based on the perceived importance of the proposed content and the perceived rigor of the proposed methodology (figure 2).

Registered Reports are currently the most effective methodological antidote to minimise the burden of publication bias and related QRPs in science.^{61 62} Registered Reports are provided an in-principle acceptance based on the perceived importance and methodological quality of the proposed study rather than on the novelty and significance of study findings. By receiving an in-principle acceptance for a Registered Report, it is plausible that researchers are less motivated to adopt QRPs that generate novel study findings to overcome publication bias. Notably, the proportion of Registered Reports reporting statistically significant results in support of study hypotheses, following data collection and analysis, is approximately 40%.^{52 63}

Similar to preregistration, a Registered Report still permits researchers to undertake exploratory research but the nature of this research (e.g., exploratory or confirmatory) is transparent. A number of sports medicine, sports physiotherapy, sport science, and sports psychology journals have adopted and currently champion the Registered Reports format for original research. A curated list of resources and journals offering the Registered Reports format is maintained by the Center for Open Science (<https://cos.io/rr/>).

Open Science Framework

The Open Science Framework (OSF, <http://osf.io/>) is an online platform that promotes open, reproducible, and collaborative research.⁶⁴ The OSF provides a centralised database where preregistered studies, supplemental materials, analysis scripts, and study datasets can be uploaded to a dedicated project page to facilitate a reproducible workflow.⁶⁵ The OSF also offers a ‘Preprints’ server that allows authors to share preprints for feedback and to gain exposure (see ‘Open Science & Research Methods Groups in Sport and Exercise Medicine and Sport Science’ below for further description of Preprints).

Direct Replication

SEM research, like many scientific fields, prioritises exploration over confirmation and does not promote a culture that values replication of exploratory research findings. No assessment evaluates the reliability of research findings as rigorously as a direct replication study.⁶⁶ Large-scale efforts to replicate the results of original studies in psychology, social science, economics, and biomedicine have demonstrated varied success, baptising a ‘replication crisis’ in the life sciences.^{67–72} Very few direct replication studies are performed in SEM.^{73 74} A direct replication study aims to independently replicate the results of an exploratory research study using identical study hypotheses, methodologies, and statistical analyses.⁷⁵ Conceptual replication studies are more common in SEM research. A conceptual replication study replicates some but not all of the methodological aspects in an original exploratory study.

Conceptual differences between the original and replication study in subpopulations studied, interventions implemented, outcomes measured and/or timelines used can substantially influence study findings and limit inferences about the original study’s replicability. As a result, direct replication is the preferred approach to investigate the credibility of an original study finding.^{68–71} Nurturing a greater replication culture in SEM will improve researcher and clinician knowledge and, more importantly, certainty about the validity and reliability of research findings that we aim to integrate into clinical practice.⁷⁶

Open Science and Research Methods Groups in Sport and Exercise Medicine and Sport Science

The Society for Transparency, Openness and Replication in Kinesiology (STORK) (<http://storkkinesiology.org/>) is a new community of researchers in SEM, sports physiotherapy, and related fields that supports researchers adopting open science practices and improving research methods.⁷⁷ STORK is the creator and home of SportRxiv, an open science server that facilitates and encourages the sharing of Preprints in SEM and Sport Sciences research. A Preprint is a complete draft of a research paper that is shared publicly before it is submitted to a journal for peer review. By posting a Preprint to a server such as SportRxiv, other researchers can provide critical feedback that advances the content and quality of the manuscript before it is subsequently submitted to a journal for peer review. Additionally, sharing a Preprint accelerates the dissemination of knowledge among the scientific community, thereby expediting the process by which

Continued

Box 1 Continued

science can advance. Preprints are rarely the final form of a research paper for most authors and Preprints frequently direct new readers to the subsequently published paper. Although not all journals have yet adapted their policies to explicitly accept manuscript submissions that have been previously posted as Preprints (<http://sherpa.ac.uk/romeo/index.php>), many journals including BJSM welcome manuscripts that have been posted as a Preprint.

BJSM 'Methods Matter' Group

The BJSM 'Methods Matter' scientific group is an international committee formed to educate the SEM community about the importance of methodological rigor when conducting clinical SEM research. The 'Methods Matter' group publish editorials and education reviews addressing issues in research methodology that influence the conduct, presentation, and interpretation of SEM research studies.⁷⁸ Topics and concepts are presented in a non-technical manner, targetingscientists, clinicians, clinician-scientists, and coaches.^{79 80} The education series aims to educate the reader to critically evaluate how methodological decisions influence the results of research studies.



Figure 2 Registered Reports workflow (reproduced with permission: <https://cos.io/rr/>).

whether each was planned prior to, or following, data collection.⁴³⁴ For example, transparently *HARKing* in the discussion section or in a referenced supplemental file allows the reader to identify the number and exploratory nature of study hypotheses that were actually tested, and how likely it is that subsequent findings arising from these hypotheses are true.³⁴ In an era where most journals permit online supplemental material, word count restriction is a weak excuse for a lack of transparency in scientific reporting.

Exploratory research facilitates novel discoveries in science and should not be discouraged. However, exploratory research findings should inform subsequent confirmatory studies using independent samples and preregistered study intentions (ie, hypotheses, methodologies, and analyses) to examine the validity, reproducibility, and replicability of exploratively identified findings. When research is highly exploratory, researchers can adopt open science practices to improve the credibility of their research and enhance the replicability of study findings.⁵⁶ In box 1, we draw attention to practical solutions and resources that promote open, transparent science to reduce the threat of QRPs.⁵⁷

Summary

Exploration enables scientific discovery, but potentially at the expense of accurate, replicable research. Multiple study hypotheses, outcome measures, and analytic strategies are often necessary to generate new research findings. However, when exploratory

research is falsely reported as confirmatory, readers cannot interpret how likely a statistically significant research finding is due to chance. We identified that only 60% of published, original SEM research studies reported a study hypothesis. Approximately 82% of published, original SEM studies reported that study hypotheses were either fully or partially supported by study results. Few study hypotheses in SEM research may be specified prior to study data collection, which can influence the reliability of study findings. Embracing a culture of open, reproducible science, which can contribute towards minimising the occurrence of QRPs and identifying deviations from *a priori* study intentions, will improve the credibility of SEM research.

Twitter Fionn Büttner @peanutbuttnr, Elaine Toomey @ElaineToomey1, Mark Roe @markroecoach and Eamonn Delahunt @EamonnDelahunt

Contributors FCB conceived the original idea, developed data extraction materials and composed the initial manuscript. FCB, ET, SMC and MR independently performed data extraction. ED arbitrated inter-rater disagreement. FCB, ET, SMC and ED provided comments on and contributed towards the revision of the final manuscript.

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

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

© Author(s) (or their employer(s)) 2020. No commercial re-use. See rights and permissions. Published by BMJ.

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/bjsports-2019-101863>).



To cite Büttner F, Toomey E, McClean S, *et al.* *Br J Sports Med* Epub ahead of print: [please include Day Month Year]. doi:10.1136/bjsports-2019-101863

Accepted 3 June 2020

Br J Sports Med 2020;0:1–7.
doi:10.1136/bjsports-2019-101863

ORCID iDs

Fionn Büttner <http://orcid.org/0000-0002-5987-3063>
Mark Roe <http://orcid.org/0000-0001-6615-2283>
Eamonn Delahunt <http://orcid.org/0000-0001-5449-5932>

REFERENCES

- MacAuley D. Sport and exercise medicine: building the foundations of a new discipline. *J Sci Med Sport* 2000;3:254–9.
- Bleakley C, MacAuley D. The quality of research in sports journals. *Br J Sports Med* 2002;36:124–5.
- Bandholm T, Henriksen M, Thorborg K. Slow down to strengthen sport and exercise medicine research. *Br J Sports Med* 2017;51:1453–4.
- Wagenmakers E-J, Wetzels R, Borsboom D, *et al.* Why PSYCHOLOGISTS must change the way they analyze their data: the case of psi: Comment on Bem (2011). *J Pers Soc Psychol* 2011;100:426–32.
- Ioannidis JPA. Why most clinical research is not useful. *PLoS Med* 2016;13:e1002049.
- Harris JD, Cvetanovich G, Erickson BJ, *et al.* Current status of evidence-based sports medicine. *Arthroscopy* 2014;30:362–71.
- Grant HM, Tjoumakaris FP, Maltenfort MG, *et al.* Levels of evidence in the clinical sports medicine literature: are we getting better over time? *Am J Sports Med* 2014;42:1738–42.
- Wagenmakers E-J, Wetzels R, Borsboom D, *et al.* An agenda for purely confirmatory research. *Perspect Psychol Sci* 2012;7:632–8.
- Jaeger RG, Halliday TR. On confirmatory versus exploratory research. *Herpetologica* 1998;54:564–6.
- Ioannidis JPA. Why most published research findings are false. *PLoS Med* 2005;2:e124.
- John LK, Loewenstein G, Prelec D. Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychol Sci* 2012;23:524–32.
- DeCoster J, Sparks EA, Sparks JC, *et al.* Opportunistic biases: their origins, effects, and an integrated solution. *Am Psychol* 2015;70:499–514.

- 13 Fanelli D. How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PLoS One* 2009;4:e5738.
- 14 Simmons JP, Nelson LD, Simonsohn U. False-Positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychol Sci* 2011;22:1359–66.
- 15 Wicherts JM, Veldkamp CLS, Augusteijn HEM, *et al.* Degrees of Freedom in Planning, Running, Analyzing, and Reporting Psychological Studies: A Checklist to Avoid *p*-Hacking. *Front Psychol* 2016;7:1832.
- 16 Ioannidis JPA. Why most discovered true associations are inflated. *Epidemiology* 2008;19:640–8.
- 17 Lim HC, Adie S, Naylor JM, *et al.* Randomised trial support for orthopaedic surgical procedures. *PLoS One* 2014;9:e96745.
- 18 Chahal J, Tomescu SS, Ravi B, *et al.* Publication of sports medicine-related randomized controlled trials registered in ClinicalTrials.gov. *Am J Sports Med* 2012;40:1970–7.
- 19 Lohmander LS, Roos EM. The evidence base for orthopaedics and sports medicine: scandalously poor in parts. *Br J Sports Med* 2016;50:564–5.
- 20 Silberzahn R, Uhlmann EL, Martin DP, *et al.* Many Analysts, one data set: making transparent how variations in analytic choices affect results. *Adv Methods Pract Psychol Sci* 2018.
- 21 Gelman A, Loken E. The garden of forking paths: Why multiple comparisons can be a problem, even when there is no “fishing expedition” or “*p*-hacking” and the research hypothesis was posited ahead of time*;17.
- 22 Héroux ME, Taylor JL, Gandeia SC. The use and abuse of transcranial magnetic stimulation to modulate corticospinal excitability in humans. *PLoS One* 2015;10:e0144151.
- 23 Patel CJ, Burford B, Ioannidis JPA. Assessment of vibration of effects due to model specification can demonstrate the instability of observational associations. *J Clin Epidemiol* 2015;68:1046–58.
- 24 Orben A, Przybylski AK. The association between adolescent well-being and digital technology use. *Nat Hum Behav* 2019;3:173–82.
- 25 Ioannidis JPA. Exposure-Wide epidemiology: revisiting Bradford Hill. *Stat Med* 2016;35:1749–62.
- 26 Mayo-Wilson E, Fusco N, Li T, *et al.* Multiple outcomes and analyses in clinical trials create challenges for interpretation and research synthesis. *J Clin Epidemiol* 2017;86:39–50.
- 27 Dodd S, White I, Williamson P. Departure from treatment protocol in published randomised controlled trials: a review. *Trials* 2011;12:A129.
- 28 Page MJ, Higgins JPT, Clayton G, *et al.* Empirical evidence of study design biases in randomized trials: systematic review of Meta-Epidemiological studies. *PLoS One* 2016;11:e0159267.
- 29 Munafò MR, Nosek BA, Bishop DVM, *et al.* A manifesto for reproducible science. *Nat Hum Behav* 2017;1:s41562–016.
- 30 Colquhoun D. An investigation of the false discovery rate and the misinterpretation of *p*-values. *R Soc Open Sci* 2014;1:140216.
- 31 Dwan K, Gamble C, Williamson PR, *et al.* Systematic review of the empirical evidence of study publication bias and outcome reporting bias - an updated review. *PLoS One* 2013;8:e66844.
- 32 Page MJ, Higgins JPT. Rethinking the assessment of risk of bias due to selective reporting: a cross-sectional study. *Syst Rev* 2016;5:108.
- 33 Kerr NL. HARKing: hypothesizing after the results are known. *Pers Soc Psychol Rev* 1998;2:196–217.
- 34 Hollenbeck JR, Harking WPM. Sharking, and Tharking: making the case for post hoc analysis of scientific data. *J Manag* 2017;43:5–18.
- 35 Rubin M. When does HARKing hurt? identifying when different types of undisclosed post hoc Hypothesizing harm scientific progress. *Review of General Psychology* 2017;21:308–20.
- 36 Banks GC, O’Boyle EH, Pollack JM, *et al.* Questions about questionable research practices in the field of management. *J Manage* 2016;42:5–20.
- 37 Motyl M, Demos AP, Carsel TS, *et al.* The state of social and personality science: rotten to the core, not so bad, getting better, or getting worse? *J Pers Soc Psychol* 2017;113:34–58.
- 38 Higginson AD, Munafò MR. Current incentives for scientists lead to Underpowered studies with erroneous conclusions. *PLoS Biol* 2016;14:e2000995.
- 39 Martinson BC, Anderson MS, de Vries R. Scientists behaving badly. *Nature* 2005;435:737–8.
- 40 Fanelli D. “Positive” results increase down the Hierarchy of the Sciences. *PLoS One* 2010;5:e10068.
- 41 Roese NJ, Vohs KD, Bias H. Hindsight bias.. *Perspect Psychol Sci* 2012;7:411–26.
- 42 Bones AK. We knew the future all along: scientific Hypothesizing is much more accurate than other forms of Precognition-A Satire in one part. *Perspect Psychol Sci* 2012;7:307–9.
- 43 Smaldino PE, McElreath R. The natural selection of bad science. *R Soc Open Sci* 2016;3:160384.
- 44 Song F, Parekh S, Hooper L, *et al.* Dissemination and publication of research findings: an updated review of related biases. *Health Technol Assess* 2010;14:1–193.
- 45 Rawat S, Meena S. Publish or perish: where are we heading? *J Res Med Sci Off J Isfahan Univ Med Sci* 2014;19:87–9.
- 46 Fanelli D. Do pressures to publish increase scientists’ bias? an empirical support from US states data. *PLoS One* 2010;5:e10271.
- 47 Edwards MA, Roy S. Academic research in the 21st century: maintaining scientific integrity in a climate of perverse incentives and Hypercompetition. *Environ Eng Sci* 2017;34:51–61.
- 48 Nosek BA, Spies JR, Motyl M. Scientific utopia: II. restructuring incentives and practices to promote truth over Publishability. *Perspect Psychol Sci* 2012;7:615–31.
- 49 Dienes Z. Understanding psychology as a science 9780230542310;185.
- 50 Sterling TD. Publication decisions and their possible effects on inferences drawn from tests of Significance-Or vice versa. *J Am Stat Assoc* 1959;54:30–4.
- 51 Sterling TD, Rosenbaum WL, Weinkam JJ. Publication decisions revisited: the effect of the outcome of statistical tests on the decision to publish and vice versa. *Am Stat* 1995;49:108–12.
- 52 Scheel AM, Schijen M, Lakens D. An excess of positive results: comparing the standard psychology literature with registered reports. *PsyArxiv* 2020.
- 53 Pulverer B, Brave Dthe. Dora the brave. *Embo J* 2015;34:1601–2.
- 54 Brembs B, Button K, Munafò M. Deep impact: unintended consequences of Journal RANK. *Front Hum Neurosci* 2013;7:291.
- 55 Brembs B. Prestigious science journals struggle to reach even average reliability. *Front Hum Neurosci* 2018;12.
- 56 Markowitz F. Five selfish reasons to work reproducibly. *Genome Biol* 2015;16:274.
- 57 Crüwell S, van Doorn J, Etz A, *et al.* 7 easy steps to open science: an annotated reading list. *PsyArxiv* 2018.
- 58 Lee H, Lamb SE, Bagg MK, *et al.* Reproducible and replicable pain research: a critical review. *Pain* 2018;159:1683–9.
- 59 Chambers CD. Registered reports: a new publishing initiative at cortex. *Cortex* 2013;49:609–10.
- 60 Hummer L, Nosek BA, Errington T. Evaluating registered reports: a naturalistic comparative study of article impact. *OSF Prepr* 2017.
- 61 Nosek BA, Lakens D. Registered reports: a method to increase the credibility of published results. *Soc Psychol* 2014;45:137–41.
- 62 Hardwicke TE, Ioannidis JPA. Mapping the universe of registered reports. *Nat Hum Behav* 2018;2:793–6.
- 63 Allen CPG, Mehler DMA. Open science challenges, benefits and tips in early career and beyond. *PsyArxiv* 2018.
- 64 Foster ED, Deardorff A. Open science framework (OSF). *J Med Libr Assoc JMLA* 2017;105:203–6.
- 65 Nosek BA, Alter G, Banks GC, *et al.* Promoting an open research culture. *Science* 2015;348:1422–5.
- 66 Kvarven A, Strømland E, Johannesson M. Comparing meta-analyses and preregistered multiple-laboratory replication projects. *Nat Hum Behav* 2020;4:423–34.
- 67 Begley CG, Ioannidis JPA. Reproducibility in science: improving the standard for basic and preclinical research. *Circ Res* 2015;116:116–26.
- 68 Camerer CF, Dreber A, Holzmeister F, *et al.* Evaluating the replicability of social science experiments in nature and science between 2010 and 2015. *Nat Hum Behav* 2018;2:637–44.
- 69 Collaboration OS. Estimating the reproducibility of psychological science. *Science* 2015;349.
- 70 Klein RA, Ratliff KA, Vianello M, *et al.* Investigating variation in Replicability. *Soc Psychol* 2014;45:142–52.
- 71 Klein RA, Vianello M, Hasselman F, *et al.* Many Labs 2: investigating variation in Replicability across samples and settings. *Adv Methods Pract Psychol Sci* 2018;1:443–90.
- 72 Young SS, Karr A, Deming KA. Deming, data and observational studies. *Signif* 2011;8:116–20.
- 73 Andersson SH, Bahr R, Clarsen B, *et al.* Risk factors for overuse shoulder injuries in a mixed-sex cohort of 329 elite handball players: previous findings could not be confirmed. *Br J Sports Med* 2017.
- 74 Tsuchida WT, Siu AM, Yamashita N, *et al.* Comparison of neuropsychological test scores of high school athletes in high and low contact sports: a replication study. *Appl Neuropsychol Child* 2018;7:14–20.
- 75 Simons DJ. The value of direct replication. *Perspect Psychol Sci* 2014;9:76–80.
- 76 Coles NA, Tiokhin L, Scheel AM, *et al.* The costs and benefits of replication studies. *Behav Brain Sci* 2018;41:e124.
- 77 Caldwell A, Vigotsky AD, Nuckolls G, *et al.* Moving sport and exercise science forward: a call for the adoption of more transparent research practices. *SportRxiv* 2019.
- 78 Nielsen Rasmus Østergaard, Shrier I, Casals M, *et al.* Statement on methods in sport injury research from the first methods matter meeting, Copenhagen, 2019. *J Orthop Sports Phys Ther* 2020;50:226–33.
- 79 Roe M, Delahun E, Büttner F. ‘Theory on relativity’: why we need to be ‘absolute’ and regulate the reporting of injury risk outcome metrics in RCTs in sport and exercise medicine (Methods Matter series). *Br J Sports Med* 2020;54:502–3.
- 80 Verhagen E, Stovitz SD, Mansournia MA, *et al.* BISM educational editorials: methods matter. *Br J Sports Med* 2018;52:1159–60.