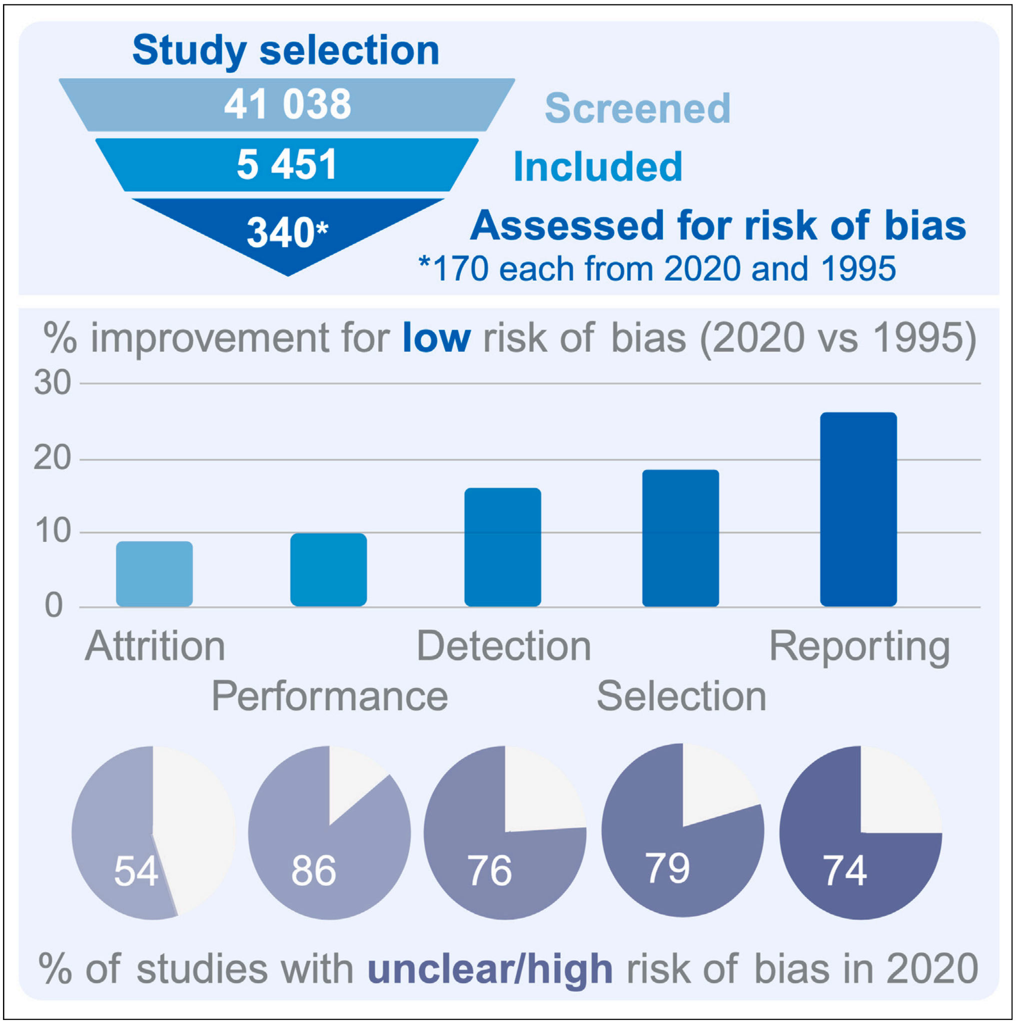
iScience

CellPress
OPEN ACCESS

Article

Risk of bias in exercise science: A systematic review of 340 studies



Nicholas Preobrazenski, Abby McCaig, Anna Turner, ..., Youssef Zaza, Stephanie Lu, Brendon J. Gurd

gurdb@queensu.ca

Highlights

Studies in exercise science have unclear or high risk of bias

The risk of bias in more recent studies is significantly reduced

There is need for rigorous reporting on experimental bias in exercise science

Preobrazenski et al., iScience 27, 109010
March 15, 2024
https://doi.org/10.1016/j.isci.2024.109010

iScience

CellPress
OPEN ACCESS

Article

Risk of bias in exercise science: A systematic review of 340 studies

Nicholas Preobrazenski,[1] Abby McCaig,[2] Anna Turner,[2] Maddy Kushner,[2] Lauren Pacitti,[2] Peter Mendolia,[2] Ben MacDonald,[2] Kristi Storoschuk,[2] Tori Bouck,[2] Youssef Zaza,[2] Stephanie Lu,[2] and Brendon J. Gurd[2,3,*]

SUMMARY

Risk of bias can contribute to irreproducible science and mislead decision making. Analyses of smaller subsections of the exercise science literature suggest many exercise science studies have unclear or high risk of bias. The current review (osf.io/jznv8) assesses whether this unclear or high risk of bias is more widespread in the exercise science literature and whether this bias has decreased since the publication of the 1996 Consolidated Standards of Reporting Trials (CONSORT) guidelines. We report significant reductions in selection, performance, detection, and reporting biases in 2020 compared with 1995 in the 340 of 5,451 studies assessed using the Cochrane Risk of Bias tool. Despite these improvements, most 2020 studies still had unclear or high risks of bias. These results underscore the need for methodological vigilance, adherence to reporting standards, and education on experimental bias. Factors contributing to these improvements, such advancements in education and journal requirements, remain uncertain.

INTRODUCTION

Background

As randomized control trials gained popularity in the 1980s, so did the need to enhance reporting quality to mitigate risk of bias.[1–7] This need was partly driven by several reports demonstrating that inadequate protections against risks of experimental bias (e.g., selection, performance, detection, attrition, reporting, etc.) produced overly favorable treatment effects.[8–10] Inadequate protections against these biases have since fueled the growing "reproducibility crisis"[11,12] and potentially undermined regulatory and clinical decision making.[13] Importantly, the absence of adequate reporting (i.e., high/unclear risk of bias)—rather than the explicit demonstration of bias—is implicated in augmented treatment effects.[14]

In 1996, the Consolidated Standards of Reporting Trials (CONSORT) guidelines[15–17] were published to provide best practices for reporting randomized controlled trials with the aim of mitigating bias and improving reproducibility. The Cochrane Collaboration Bias Assessment Tool[18] was subsequently published as a means of assessing reporting quality and risk of bias. Although the CONSORT guidelines improved reporting quality of randomized controlled trials in the clinical literature,[19–21] meta-epidemiological studies of papers published in the post-CONSORT era judged that most studies continued to have a high or unclear risk of bias across several other domains of science[14,22–25] including clinical[26] and preclinical[27] literature. These poor reporting practices persist despite multiple publications detailing the economic[28] and scientific ramifications[11,29–33] of failing to reduce bias and improve reproducibility.

Exercise science appears to suffer from similarly poor reporting practices[34] despite the endorsement of bias-reducing practices by multiple exercise science journals. Indeed, we[35–37] and several other human researchers (Fanchini M. et al. and Guevara S.A. et al.[38,39]) have recently reported high or unclear risk of bias in small subsections of the aerobic and strength and conditioning literature. Whether the risk of bias is more widespread in the exercise science literature and whether this bias has decreased since the publication of the CONSORT guidelines in 1996 remains unexamined. Filling these knowledge gaps will help exercise science researchers appreciate risk of bias in past studies and improve awareness of experimental best practices that will help to mitigate risk of bias in future work.

The purpose of this systematic review was to test two hypotheses: (1) most exercise papers published in 2020 have a high or unclear risk of bias across five types of bias (e.g., selection, performance, detection, attrition, and reporting); and (2) the 1996 CONSORT guidelines contributed to a reduced proportion of exercise science studies judged to have high or unclear risk of bias in studies published in 2020 compared with studies published in 1995.

[1]Faculty of Medicine, University of Ottawa, Ottawa, ON K1H 8M5, Canada
[2]School of Kinesiology and Health Studies, Queen's University, Kingston, ON K7L 3N6, Canada
[3]Lead contact
*Correspondence: gurdb@queensu.ca
https://doi.org/10.1016/j.isci.2024.109010

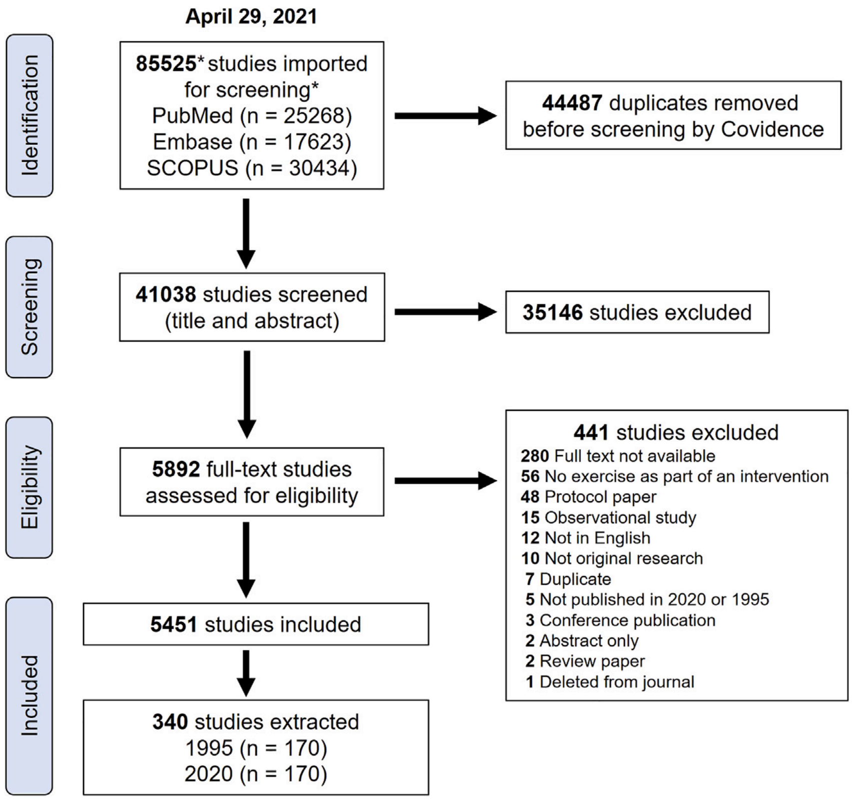


Figure 1. Flow diagram of the study selection process

*Total number is greater than individual searches due to SCOPUS only allowing extraction of the first 2000 studies per search; redundant extractions were required and led to more than 30434 studies imported into Covidence.

Methods

This systematic review adhered to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses checklist[40] (see Data S1-sheet S1). Abstract/title screening, full-text review, and data extraction were conducted using Covidence's systematic review software (Veritas Health Innovation, Australia). The intended methods were documented on Open Science Framework (osf.io/jznv8) on April 29, 2021, prior to the literature searches, data extraction, and data analysis. The registration was updated on July 8, 2022, to include a finalized analysis plan.

Eligibility criteria

Studies were included if they satisfied all the following inclusion criteria: they (1) involved humans or animals, (2) were an original, published research study, (3) incorporated exercise as part of an intervention, (4) were in English, (5) were published in 1995 or 2020, and (6) could be sourced from two different libraries, repeated online searches, or an attempt to contact its authors. Studies were excluded if they: (1) reported previously published data, (2) were a conference publication, or (3) were observational studies.

Search strategy

We conducted literature searches of studies published in 1995 and 2020 in PubMed, EMBASE, and Scopus on April 29, 2021, following protocol registration (osf.io/jznv8). Search strategies used "exercise" as the key concept. All search strategies attempted to exclude non-English and non-primary research (see Data S1-sheet S2). Titles and abstracts were then imported into Covidence from the database searches on April 30, 2021, and Covidence automatically removed duplicates. Reference lists of included studies were not searched.

Selection process

Study selection followed a two-step process and was independently completed by two reviewers. Both reviewers met to resolve conflicts, and a third reviewer resolved the remaining disagreements. First, titles and abstracts were screened to identify studies appearing to meet eligibility criteria. Second, full texts were downloaded for studies that passed title and abstract screening. Studies removed during full-text screening were assigned a reason for exclusion (see Figure 1 and Data S1-sheet S3).

Data extraction and risk of bias assessment

After conducting full-text screening, two authors independently extracted data from the included studies. A third reviewer resolved any conflicts. Extracted outcome data and data items are provided in Data S1-sheet S4. Sex and gender were interchanged (i.e., "woman" with "female" and "man" with "male") for data extraction purposes.

The risk of bias in included studies was assessed using the Cochrane Collaboration Risk of Bias Assessment Tool[18] by at least two people working independently. A third reviewer resolved any conflicts. The risk for each source of bias in these studies was judged as either "high", "low", or "unclear". Outcome-specific judgments were completed for outcome assessor bias and attrition bias; however, primary outcomes were rarely reported so reviewers often used their best judgment based on the study title and results highlighted in the abstract. Studies that reported an adequate methodology to protect against a given source of bias either within the text or in a protocol registration document (e.g., blinding outcome assessors to protect against detection bias) were judged as having a "low" risk of bias. Studies that did not report information regarding a given methodology were judged as having an "unclear" risk of bias. Studies that reported being "open-label" solely in a registration document were judged as having a "high" risk of performance and detection bias. Studies were judged as having a "high" risk of reporting bias if they did not report publicly registering their trial, and all studies published in 1995 were judged to have a high risk of reporting bias because these studies predated trial registries. When available, study protocol documents were used to inform judgments in cases of unclear or a lack of reporting in the published manuscript.

Data synthesis and statistical analysis

Descriptive statistics were used for domain, type of exercise, type of training, and categories of bias in Microsoft excel (v.16.72). For selection bias, we used the higher number of unclear/high judgments from its constituents (sequence generation and allocation concealment). Although our initial intention was to assess risk of bias in all included studies, we modified our assessment approach after the full-text review to assess 170 studies published in 1995 and 170 studies published in 2020 (modification was included in an update to our study registration on July 8, 2022). This sample size was selected because a sample size of 167 provided 80% power to detect a 15% difference in proportions between years with a 95% confidence level using the formula from section 3.1 in Wang and Chow.[41] We used lower-than-expected proportions of unclear/high studies (e.g., 65% vs. 50%; Cohen's $h = 0.3$) in our power calculation because this yielded a more conservative sample size estimate compared with using a 15% difference at higher population proportions (see Table 6.2.1 in Cohen[42]).

Two-tailed Z score tests for two population proportions with no continuity correction[43] were performed to determine whether the proportions of studies judged to have unclear or high risks of bias were different between studies published in 1995 and 2020. The Bonferroni correction was applied to account for multiple comparisons and avoid spurious positives (Bonferroni corrected $\alpha = 0.05/5$ types of bias $= 0.01$). Significance was therefore set at $p < 0.01$.

RESULTS

Study selection

Figure 1 presents a flow diagram of the study selection process. The literature search retrieved 85,525 studies, and Covidence removed 44,487 duplicates. Out of the 41,038 studies entering title and abstract screening, 35,146 were deemed irrelevant and were subsequently excluded. Of the 5,892 full texts downloaded, 441 studies were excluded, leaving 5,451 studies included for review. Covidence listed these studies in alphabetical order according to the first author's last name. From this list, we extracted the first 170 studies appearing from 1995 and the first 170 studies appearing from 2020. Reference information for all extracted studies can be found in Data S1-sheet S4.

Study characteristics

Figure 2 displays a combined and between-year comparison of characteristics of included studies. Data S1-sheet S5 presents study characteristics by exercise domain, modality, type, and population and graphically presents combined and year-based study characteristics. Most (78%) studies involved humans (1995: n = 127; 2020: n = 137). Biochemistry [n = 139 (41%), 1995: n = 79; 2020: n = 60], aerobic [n = 213 (63%), 1995: n = 134; 2020: n = 79], and exercise training [n = 240 (71%), 1995: n = 98, 2020: n = 142] represented the most common domain, exercise modality, and type of exercise, respectively. Nearly half of the extracted studies (n = 146, 48%) were open access at time of data extraction (1995: n = 47; 2020: n = 117). Other study characteristics (e.g., sample sizes, study durations, etc.) can be found in Data S1-sheet S4. Data S1-sheet S5 presents percentage of studies reporting sample size calculations by exercise modality, type, and population.

Risk of bias

Table 1 displays a summary of the risk of bias judgments for the 340 extracted studies. Figure 3 illustrates both the proportion of these studies that were categorized as high or unclear risk of bias and the variation in this proportion from 1995 to 2020. The percent of all studies with high and unclear risk of bias and the change in the percent of studies with high and unclear risk of bias between 1995 and 2020 are presented in Figure 3. Data S1-sheet S4 presents study-level judgments, and Data S1-sheet S5 presents risk of bias by exercise modality, type, and population. In general, we observed unclear or high risk of bias among studies that involved exercise as part of an intervention. Encouragingly, the studies published in 2020 had a significantly lower proportion of judgments with unclear/high risk of selection ($p < 0.00001$; Cohen's $h = 0.65$; proportions' difference between 1995 and 2020 with 95% confidence intervals: 18.8% [12.2, 25.4]), performance ($p = 0.00135$; $h = 0.36$; 10.0% [3.98, 16.0]), detection ($p < 0.0001$; $h = 0.45$; 15.9% [8.26, 23.5]), and reporting ($p < 0.00001$; $h = 1.07$; 25.9% [19.3, 32.5]) bias. There was no

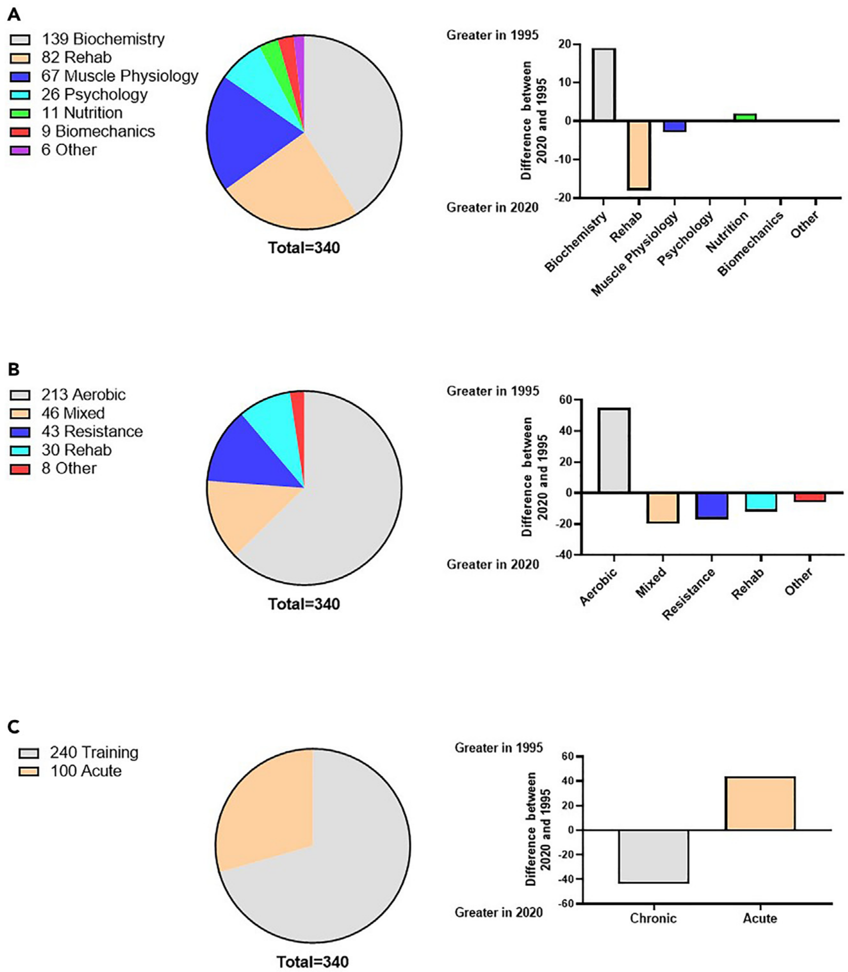


Figure 2. Composition of articles included in analysis and change in risk of bias from 1995 to 2020
Composition of (A) domain, (B) exercise modality, and (C) exercise type between studies published in 1995 and 2020. N = 340 studies.

difference between 1995 and 2020 studies for attrition bias ($p = 0.079$; $h = 0.19$; 9.4% [1.04, 19.8]). Only seven studies (~2%) were judged to have low risks of all sources of bias (selection, performance, detection, attrition, and reporting bias). Only 58 studies (~17%) reported performing a sample size calculation (1995: n = 5; 2020: n = 53).

Because we observed large changes in the modality, type, and population from 1995 to 2020 (Figure 2), we performed several exploratory analyses to better understand the impact of these altered proportions on our observations of improved bias. The results of these exploratory analyses are presented in Figure 4 and Data S1-sheets S6–S8. Although we did observe some significant changes in high/unclear risk of bias (e.g., within aerobic, training, and human studies; Figure 4), the relative lack of significance should be interpreted with caution given our conservative p value ($p < 0.0083$ due to six comparisons) and the small sample sizes included for several subgroups (e.g., resistance training studies, 1995: n = 14; 2020: n = 29). Given these exploratory analyses were not powered to detect smaller effects (e.g., a Cohen's h of 0.2),[42] the magnitude and direction of calculated effect sizes could be inaccurate.

DISCUSSION

This systematic review tested the hypotheses that: (1) most exercise papers published in 2020 would have a high or unclear risk of bias across five types of bias (e.g., selection, performance, detection, attrition, and reporting); and (2) the 1996 CONSORT guidelines would decrease the proportion of exercise science studies judged to have high or unclear risk of bias. The obtained results are consistent with both stated hypotheses.

The large number of studies in this review with unclear or high risks of selection, performance, detection, attrition, and reporting bias carries substantial implications[28,33,44] (see section —"*Understanding and mitigating bias in exercise science: Types and strategies*"—for a detailed discussion). Overall unclear or high risk of bias in hundreds of studies across the full range of exercise sciences (Table 1 and Figure 2;

Table 1. Risk of bias in studies involving exercise published in 1995 or 2020 (N = 340)

	Selection Bias		Performance Bias	Detection Bias	Attrition Bias	Reporting Bias
	Random sequence generation	Allocation concealment	Blinded participants and personnel	Blinded outcome assessment	Incomplete outcome data	Selective reporting
Overall (N = 340)						
# unclear or high (%)	268 (79)	300 (88)	309 (91)	285 (84)	198 (58)	296 (87)
A 1995 (n = 170)						
# unclear or high (%)	150 (88)	166 (98)	163 (96)	156 (92)	107 (63)	170 (100)
a2020 (n = 170)						
# unclear or high (%)	118 (69)	134 (79)	146 (86)	129 (76)	91 (54)	126 (74)
1995 → 2020	↓ 19%[a]	↓ 19%[a]	↓ 10%[a]	↓ 16%[a]	↓ 9%	↓ 26%[a]

[a]Statistically significant difference at $p < 0.01$.

Data S1-sheet S4) aligns with recent reviews in subsections of the human exercise science literature.[35–37,45] Our results underscore a need for improved methodological rigor and transparency in study design, execution, and reporting to enhance the reliability and applicability of exercise science research. This means researchers should consider including and reporting robust randomization and blinding procedures, standardized outcome measures, proactive attrition management, and comprehensive result reporting.

Elevating standards in exercise science with CONSORT and Cochrane

Despite the substantial proportion of 2020 studies judged to have high or unclear risk of bias, there were marked improvements in risk of bias between 1995 and 2020 (Figure 3). These findings align with previous work examining changes in risk of bias in articles published in medical journals from 2011 to 2014.[26] Moreover, while sample size calculations remained relatively infrequent in 2020 (53/170 studies), this represents a considerable improvement from 1995 (5/170). These improvements are encouraging given the impact that high-quality reporting can have on reproducibility, reliability, and trust.[46–49] Although the exact mechanisms underlying the improving reporting practices within exercise science are unclear, it is likely that the production of numerous robust reporting guidelines in the past 25 years has contributed. These guidelines have been developed to help improve the reporting of research findings and include CONSORT ,[50] PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses),[40] SPIRIT (Standard Protocol Items: Recommendations for Interventional Trials),[51] TIDieR (Template for Intervention Description and Replication),[52] and CERT (Consensus-based Clinical Case Reporting).[53] The importance of reporting has been highlighted in the biomedical[54] and exercise sciences,[55] and many exercise science researchers have advocated for more transparent research practices.[35,37,47,48,56–60]

The CONSORT statement[50]—a standardized 25-item checklist designed to ensure comprehensive and transparent reporting of all relevant information—has positively impacted randomized controlled trial reporting quality.[19–21] There is significant overlap between the CONSORT guidelines and Cochrane's risk of bias assessment tool,[18] particularly in the areas of blinding, incomplete outcome data, and selective reporting. The CONSORT framework insists on disclosing whether blinding was implemented, and if so, the identities of those blinded. This clear reporting is crucial for Cochrane's risk of bias assessment tool to assess the risk of bias related to blinding of participants, study personnel, and outcome assessors. For incomplete outcome data, CONSORT prompts the reporting of the number and reasons for participant dropouts. Cochrane's tool then assesses the risk of bias related to incomplete outcome data, helping to determine if the dropout rate was high and if systematic differences existed between groups. Furthermore, CONSORT also advocates for the reporting of all outcomes, inclusive of negative results, which then helps researchers use Cochrane's tool when evaluating the risk of reporting bias. Although we posit that the CONSORT guidelines have positively impacted reporting practices in exercise science between 1995 and 2020 given their publication in 1996, it is likely that improved reporting practices have been impacted by a myriad of factors.

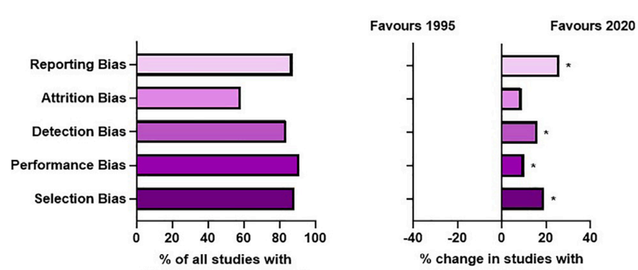


Figure 3. The percent change of 340 studies with high/unclear risk of bias between studies published in 1995 (n = 170) and 2020 (n = 170)

* Statistically significant at $p < 0.01$.

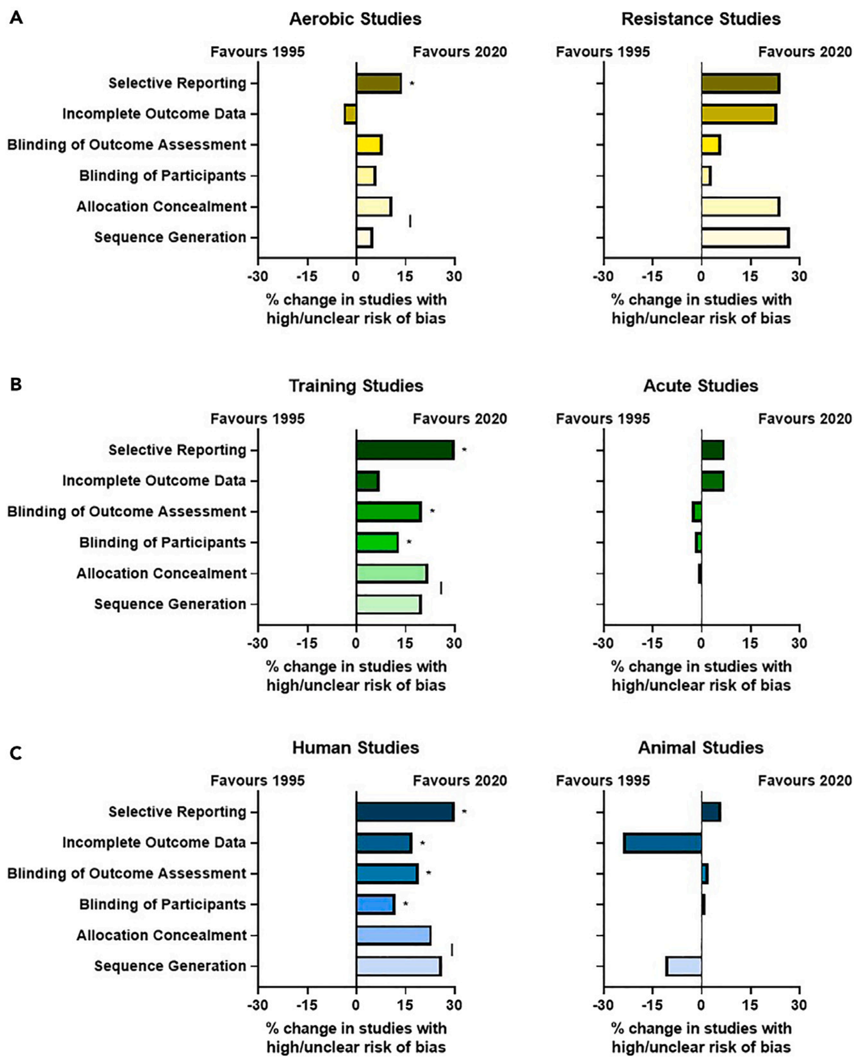


Figure 4. Exploratory analyses presenting differences in percentage of studies judged to have high/unclear risk of bias in

(A) Aerobic (1995: n = 134; 2020: n = 78) and resistance (1995: n = 14; 2020: n = 29) studies, (B) Training (1995: n = 98; 2020: n = 142) and acute (1995: n = 72; 2020: n = 28) studies, and (C) Human (1995: n = 127; 2020: n = 137) and animal (1995: n = 43; 2020: n = 33) studies. Note that sequence generation/allocation concealment equate to selection bias, blinding of participants to performance bias, blinding of outcome assessment to detection bias, incomplete outcome data to attrition bias, and selective reporting to reporting bias. * Statistically significant at $p < 0.0083$. | Statistically significant change in selection bias at $p < 0.0083$.

Navigating the nuances of bias in exercise science with exploratory analyses

The exploratory analyses showcased in Figure 4 and Data S1-sheets S6–S8, while intriguing, warrant careful interpretation. We did not conduct power analyses to detect differences in proportions across exercise type, modality, or populations. These findings are presented to facilitate future hypotheses rather than to test *a priori* hypotheses. For example, resistance exercise studies appear to have reduced bias more than aerobic exercise studies since 1995. However, this observation could be attributed to the fewer number of resistance exercise studies (43 resistance vs. 212 aerobic) and/or a higher relative proportion (37% in 1995 vs. 67% in 2020). The apparent lack of change in acute exercise studies and those involving animals might be explained by similar reasoning or even by differences in journal impact factor.[26,61] To appropriately determine differences in risk of bias by exercise type, modality, or population, comparisons with adequate power are necessary.

Understanding and mitigating bias in exercise science: Types and strategies

Selection bias

Selection bias occurs when participants are preferentially assigned to one group over another,[62] leading to an unrepresentative sample of the intended population. For instance, if fitter individuals are predominantly assigned to the intervention group, and less fit individuals are assigned to the control group, outcome differences could be attributed to these initial disparities rather than the intervention.

Selection bias hinges on two factors: random sequence generation and allocation concealment. Sequence generation refers to the randomization process of assigning participants to different groups. Flaws in this process can spawn selection bias and can lead to systematic baseline differences between groups. If sequence generation lacks true randomness, researchers could anticipate group assignments and manipulate allocation to achieve a desired outcome. Allocation concealment ensures the allocator remains blind to treatment assignments until after a participant enrolls. Inadequate concealment could introduce selection bias as knowledge of group assignment might sway the allocator, causing systematic differences that can compromise internal validity.

Neglecting to address selection bias could lead to skewed study results and lead to misguided conclusions and inappropriate exercise interventions. To mitigate selection bias, researchers should employ randomized allocation methods, such as computer-generated randomization and blind investigators to allocation sequences.[63–65] The observation that most studies in the current review demonstrated an unclear or high risk of selection bias demonstrates a clear need to improve rigor, or at least reporting, of participant randomization processes in exercise science.

Performance bias

Performance bias arises from systematic differences in the care provided to different groups or from behavioral changes by participants or researchers due to awareness of the assigned interventions.[18] This can distort the true effects of an intervention and mislead conclusions. Blinding, where possible, is instrumental in mitigating performance bias. This involves blinding participants, personnel, and outcome assessors, ensuring impartial implementation and evaluation of interventions. Our analysis found a high proportion of studies at risk of performance bias (Table 1), which likely reflects the intrinsic difficulties of blinding participants and personnel in exercise studies.[66,67] These challenges underscore a need for creative solutions regarding blinding in exercise trials. Possible solutions could include the use of sham or placebo controls where appropriate, comprehensive training for researchers to ensure equal treatment across participant groups, and objective outcome measures.

Detection bias

Detection bias stems from systematic differences in how outcomes are determined based on knowledge of participants' group assignments. This bias can distort the true impact of an intervention by overestimating or underestimating its effects.[68] Countermeasures against detection bias include blinding outcome assessors to the intervention groups and using objective, standardized outcome measures. Most studies in our systematic review had an unclear risk of detection bias (Table 1) because they did not report whether outcome assessors were blinded. Strategies to blind outcome assessors have been discussed in our previous work.[35,69] Ensuring thorough reporting on blinding procedures, or justifying their absence, can help enhance the reliability of results and maintain confidence in the outcomes of exercise science studies.

Attrition bias

Attrition bias occurs when participants drop out of the study or are lost to follow-up, and the reasons for dropout are related to the outcome being measured.[70] This can lead to a significant reduction in statistical power and potential skewing of results, as the remaining participants may not accurately represent the initial population. To mitigate the risk of attrition bias, particularly in instances of high dropout rates are high or systematic variance between groups, an intention-to-treat analysis should be performed.[17] This approach, which includes all participants in the final analysis according to their original group assignment regardless of their adherence or dropout, can help maintain the original balance achieved through randomization and mitigate potential attrition bias.

Reporting bias

Reporting bias occurs when the dissemination of research findings is influenced by the nature and direction of results.[71] This can lead to an overrepresentation of positive results, creating an inflated perception of effectiveness. To combat reporting bias, researchers should rigorously adhere to guidelines such as the CONSORT statement, which ensures comprehensive reporting of both positive and negative outcomes. Additionally, pre-registering trials (e.g., via Open Science Framework or ClinicalTrials.gov) and committing to publish results regardless of outcome are practices to safeguard against reporting bias.[72] Considering that ClinicalTrials.gov was launched in 2000 and Open Science Framework in 2012, studies published in 1995 in our review were assessed as having an unclear/high risk of bias.

The need for sample size calculations in exercise science

Statistical power analysis is essential for determining the sample size needed to detect a given effect size. The term "power" refers to the likelihood of a test correctly identifying a true effect. Thus, studies lacking adequate power may fail to detect an effect that is truly present, leading to a false negative, or a Type II error and nonreplicable results.[73] Such errors can culminate in preventable costs, unethical consequences, and irreproducible results.

Many reports[35,45,55,56,58,59,74] have highlighted the need for more studies in the sport and exercise sciences to perform *a priori* power calculations to reduce the risk of Type II errors.[75] Although the issue of low statistical power in the exercise sciences has been known for many decades,[76,77] the rise in studies reporting power calculations from a meager 2.8% in 1995 to 31.2% in 2020 is substantial but still insufficient.

Limitations of the study

While this review includes a notably large number of studies for an exercise science review (340 studies; 17,584 participants), it is important to acknowledge its limitations. First, the Cochrane's Risk of Bias 1.0 Tool may not have captured nuances that could have been uncovered by using the more comprehensive and demanding[78] Cochrane's Risk of Bias 2.0 Tool.[79] However, we selected the 1.0 Tool to facilitate comparisons with other studies that have performed bias assessments in the preclinical[27] and clinical[26,35] sciences.

Second, without a comprehensive evaluation of results reporting, we cannot definitively state whether the proportion of studies with unclear or high risk of bias has decreased since 1995. Although "unclear" judgments can reflect poor reporting,[18] the absence of comprehensive reporting does not necessarily indicate the presence of bias or poor methods.[80,81] While the CONSORT checklist could be considered for assessing reporting quality using items linked to risk of bias, it was not designed for quality assessment.

Third, ambiguous reporting of study design and randomization procedures may have resulted in potentially inaccurate assessment(s) because risk of bias assessment tools is generally tailored for particular study types and designs.[82]

Fourth, we cannot conclude the 1996 CONSORT guidelines *per se* reduced risk of bias because only 340 of 5,451 potential studies were assessed and unmeasured influences (e.g., improved graduate training in research methodology, opportunities for feedback, etc.) ostensibly contributed to reduced risks of bias.

Lastly, the unequal composition of domains, exercise modalities, exercise types, and populations (human and animal) between the 1995 and 2020 studies included in this review may have led to inaccurate results. For instance, the higher proportion of exercise training studies in 2020 may have exaggerated the decrease in risk of bias between 1995 and 2020 studies because training studies appeared more likely to protect against biases. This result could mean the decrease in risk of bias from 1995 to 2020 was due to an inadvertently higher proportion of training studies in 2020 rather than an actual decrease in risk of bias over time. Moreover, we could not employ stratified sampling because doing so would require extracting data from all 5,451 studies. The lack of stratified sampling and searching reference lists of included studies introduces a potential selection bias because our sample of studies may not reflect a representative sample of the 5,451 available studies. To enhance the robustness of our findings, future analyses could concentrate on a specific area of exercise science, incorporate stratified sampling, and directly compare risk of bias in larger and balanced samples of studies grouped by the categories presented in the current review.

Conclusion

Most exercise papers published in 2020 had a high or unclear risk of bias across five types of bias (e.g., selection, performance, detection, attrition, and reporting), and results demonstrated a reduced risk of bias in a large set of exercise science studies since the publication of the 1996 CONSORT guidelines. Despite these improvements, there is a continued need for improved random sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessment, incomplete outcome data, and selective reporting in exercise science research. These findings underscore a need for increased methodological vigilance, adherence to rigorous reporting standards, and teaching about experimental bias in educational curricula. We can hopefully increase the validity and reliability of exercise science research by continuously guarding against sources of bias. This, in turn, may bolster the credibility of our research, inform evidence-based clinical practices, and lead to better health outcomes in applied settings.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Material availability
  - Data and code availability
- METHOD DETAILS
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2024.109010.

AUTHOR CONTRIBUTIONS

N.P. conceived of the study, wrote the original draft of the manuscript, acquired, analyzed, and interpreted data, supervised, and engaged in project administration. B.G. assisted with study design, manuscript editing, made substantial contributions to the analysis and interpretation of data, supervised, and engaged in project administration. Other authors contributed to the acquisition and interpretation of data, and they revised it critically for important intellectual content. All authors approved the version to be published, approved the order of authors, and agree to be accountable for all aspects of the work.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: October 4, 2023
Revised: December 19, 2023
Accepted: January 22, 2024
Published: January 26, 2024

REFERENCES

1. Pocock, S.J., Hughes, M.D., and Lee, R.J. (1987). Statistical problems in the reporting of clinical trials. N. Engl. J. Med. *317*, 426–432. https://doi.org/10.1056/NEJM198708133170706.
2. Altman, D.G. (1994). The scandal of poor medical research. BMJ *308*, 283–284. https://doi.org/10.1136/bmj.308.6924.283.
3. Gøtzsche, P.C. (1989). Methodology and overt and hidden bias in reports of 196 double-blind trials of nonsteroidal antiinflammatory drugs in rheumatoid arthritis. Control. Clin. Trials *10*, 31–56. https://doi.org/10.1016/0197-2456(89)90017-2.
4. Freiman, J.A., Chalmers, T.C., Smith, H., and Kuebler, R.R. (1978). The importance of beta, the type ii error and sample size in the design and interpretation of the randomized control trial. N. Engl. J. Med. *299*, 690–694. https://doi.org/10.1056/NEJM197809282991304.
5. Bland, J.M., Jones, D.R., Bennett, S., Cook, D.G., Haines, A.P., and MacFarlane, A.J. (1985). Is the clinical trial evidence about new drugs statistically adequate? Br. J. Clin. Pharmacol. *19*, 155–160. https://doi.org/10.1111/j.1365-2125.1985.tb02626.x.
6. DerSimonian, R., Charette, L.J., McPeek, B., and Mosteller, F. (1982). Reporting on methods in clinical trials. N. Engl. J. Med. *306*, 1332–1337. https://doi.org/10.1056/NEJM198206033062204.
7. Moher, D., Dulberg, C.S., and Wells, G.A. (1994). Statistical Power, sample size, and their reporting in randomized controlled trials. J. Am. Med. Assoc. *272*, 122–124. https://doi.org/10.1001/jama.1994.03520020048013.
8. Schulz, K.F., Chalmers, I., Hayes, R.J., and Altman, D.G. (1995). Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA *273*, 408–412. https://doi.org/10.1001/jama.273.5.408.
9. Moher, D., Pham, B., Jones, A., Cook, D.J., Jadad, A.R., Moher, M., Tugwell, P., and Klassen, T.P. (1998). Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? Lancet *352*, 609–613. https://doi.org/10.1016/S0140-6736(98)01085-X.
10. Chalmers, T.C., Celano, P., Sacks, H.S., and Smith, H. (1983). Bias in treatment assignment in controlled clinical trials. N. Engl. J. Med. *309*, 1358–1361. https://doi.org/10.1056/NEJM198312013092204.
11. 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., and Ioannidis, J.P.A. (2017). A manifesto for reproducible science. Nat. Hum. Behav. *1*, 0021. https://doi.org/10.1038/s41562-016-0021.
12. Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. Nature *533*, 452–454. https://doi.org/10.1038/533452a.
13. Wallach, J.D., Gonsalves, G.S., and Ross, J.S. (2018). Research, regulatory, and clinical decision-making: the importance of scientific integrity. J. Clin. Epidemiol. *93*, 88–93. https://doi.org/10.1016/j.jclinepi.2017.08.021.
14. Page, M.J., Higgins, J.P.T., Clayton, G., Sterne, J.A.C., Hróbjartsson, A., and Savović, J. (2016). Empirical evidence of study design biases in randomized trials: systematic review of meta-epidemiological studies. PLoS One *11*, e0159267. https://doi.org/10.1371/journal.pone.0159267.
15. Begg, C., Cho, M., Eastwood, S., Horton, R., Moher, D., Olkin, I., Pitkin, R., Rennie, D., Schulz, K.F., Simel, D., and Stroup, D.F. (1996). Improving the quality of reporting of randomized controlled trials: The CONSORT statement. J. Am. Med. Assoc. *276*, 637–639. https://doi.org/10.1001/jama.276.8.637.
16. Altman, D.G., Schulz, K.F., Moher, D., Egger, M., Davidoff, F., Elbourne, D., Gøtzsche, P.C., and Lang, T.; CONSORT GROUP Consolidated Standards of Reporting Trials (2001). The revised CONSORT statement for reporting randomized trials: explanation and elaboration. Ann. Intern. Med. *134*, 663–694. https://doi.org/10.7326/0003-4819-134-8-200104170-00012.
17. Moher, D., Hopewell, S., Schulz, K.F., Montori, V., Gøtzsche, P.C., Devereaux, P.J., Elbourne, D., Egger, M., and Altman, D.G. (2010). CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. BMJ *340*, c869. https://doi.org/10.1136/bmj.c869.
18. Higgins, J.P.T., Altman, D.G., Gøtzsche, P.C., Jüni, P., Moher, D., Oxman, A.D., Savovic, J., Schulz, K.F., Weeks, L., Sterne, J.A.C., et al. (2011). The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ *343*, d5928. https://doi.org/10.1136/bmj.d5928.
19. Moher, D., Jones, A., and Lepage, L.; CONSORT Group Consolidated Standards for Reporting of Trials (2001). Use of the CONSORT statement and quality of reports of randomized trials. J. Am. Med. Assoc. *285*, 1992–1995. https://doi.org/10.1001/jama.285.15.1992.
20. Egger, M., Jüni, P., and Bartlett, C.; CONSORT Group Consolidated Standards of Reporting of Trials (2001). Value of flow diagrams in reports of randomized controlled trials. J. Am. Med. Assoc. *285*, 1996–1999. https://doi.org/10.1001/jama.285.15.1996.
21. Plint, A.C., Moher, D., Morrison, A., Schulz, K., Altman, D.G., Hill, C., and Gaboury, I. (2006). Does the CONSORT checklist improve the quality of reports of randomised controlled trials? A systematic review. Med. J. Aust. *185*, 263–267. https://doi.org/10.5694/j.1326-5377.2006.tb00557.x.
22. Wood, L., Egger, M., Gluud, L.L., Schulz, K.F., Jüni, P., Altman, D.G., Gluud, C., Martin, R.M., Wood, A.J.G., and Sterne, J.A.C. (2008). Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. BMJ *336*, 601–605. https://doi.org/10.1136/bmj.39465.451748.AD.
23. Pildal, J., Hróbjartsson, A., Jørgensen, K.J., Hilden, J., Altman, D.G., and Gøtzsche, P.C. (2007). Impact of allocation concealment on conclusions drawn from meta-analyses of randomized trials. Int. J. Epidemiol. *36*, 847–857. https://doi.org/10.1093/ije/dym087.
24. Nüesch, E., Reichenbach, S., Trelle, S., Rutjes, A.W.S., Liewald, K., Sterchi, R., Altman, D.G., and Jüni, P. (2009). The importance of allocation concealment and patient blinding in osteoarthritis trials: a meta-epidemiologic study. Arthritis Rheum. *61*, 1633–1641. https://doi.org/10.1002/art.24894.
25. Armijo-Olivo, S., Saltaji, H., Da Costa, B.R., Fuentes, J., Ha, C., and Cummings, G.G. (2015). What is the influence of randomization sequence generation and allocation concealment on treatment effects of physical therapy trials? A meta-epidemiological study. BMJ Open *5*, 1–11. https://doi.org/10.1136/bmjopen-2015-008562.
26. Dechartres, A., Trinquart, L., Atal, I., Moher, D., Dickersin, K., Boutron, I., Perrodeau, E., Altman, D.G., and Ravaud, P. (2017). Evolution of poor reporting and inadequate methods over time in 20 920 randomised controlled trials included in Cochrane reviews: research on research study. BMJ *357*, j2490. https://doi.org/10.1136/bmj.j2490.
27. Ramirez, F.D., Motazedian, P., Jung, R.G., Di Santo, P., Macdonald, Z.D., Moreland, R., Simard, T., Clancy, A.A., Russo, J.J., Welch, V.A., et al. (2017). Methodological rigor in preclinical cardiovascular studies. Circ. Res. *120*, 1916–1926. https://doi.org/10.1161/CIRCRESAHA.117.310628.
28. Freedman, L.P., Cockburn, I.M., and Simcoe, T.S. (2015). The economics of reproducibility in preclinical research. PLoS Biol. *13*, e1002165. https://doi.org/10.1371/journal.pbio.1002165.
29. Begley, C.G., and Ioannidis, J.P.A. (2015). Reproducibility in science: Improving the standard for basic and preclinical research. Circ. Res. *116*, 116–126. https://doi.org/10.1161/CIRCRESAHA.114.303819.
30. Ioannidis, J.P.A. (2005). Why most published research findings are false. PLoS Med. *2*, e124. https://doi.org/10.1371/journal.pmed.0020124.
31. Ioannidis, J.P.A. (2014). How to make more published research true. PLoS Med. *11*, e1001747. https://doi.org/10.1371/journal.pmed.1001747.
32. Freedman, L.P., and Gibson, M.C. (2015). The impact of preclinical irreproducibility on drug development. Clin. Pharmacol. Ther. *97*, 16–18. https://doi.org/10.1002/cpt.9.
33. Moher, D., Glasziou, P., Chalmers, I., Nasser, M., Bossuyt, P.M.M., Korevaar, D.A., Graham, I.D., Ravaud, P., and Boutron, I. (2016). Increasing value and reducing waste in biomedical research: Who's listening? Lancet *387*, 1573–1586. https://doi.org/10.1016/S0140-6736(15)00307-4.

34. Murphy, J., Mesquida, C., and Warne, J. (2023). Survey on the attitudes towards and perception of reproducibility and replicability in sports and exercise science. Commun. Kinesiol. *1*. https://doi.org/10.51224/cik.2023.53.
35. Bonafiglia, J.T., Islam, H., Preobrazenski, N., and Gurd, B.J. (2022). Risk of bias and reporting practices in studies comparing VO_2max responses to sprint interval vs. continuous training: a systematic review and meta-analysis. J. Sport Health Sci. *11*, 552–566. https://doi.org/10.1016/j.jshs.2021.03.005.
36. Bonafiglia, J.T., Preobrazenski, N., and Gurd, B.J. (2021). A systematic review examining the approaches used to estimate interindividual differences in trainability and classify individual responses to exercise training. Front. Physiol. *12*, 665044. https://doi.org/10.3389/fphys.2021.665044.
37. Preobrazenski, N., Islam, H., and Gurd, B.J. (2021). Molecular regulation of skeletal muscle mitochondrial biogenesis following blood flow-restricted aerobic exercise: a call to action. Eur. J. Appl. Physiol. *121*, 1835–1847. https://doi.org/10.1007/s00421-021-04669-6.
38. Fanchini, M., Steendahl, I.B., Impellizzeri, F.M., Pruna, R., Dupont, G., Coutts, A.J., Meyer, T., and McCall, A. (2020). Exercise-based strategies to prevent muscle injury in elite footballers: a systematic review and best evidence synthesis. Sports Med. *50*, 1653–1666. https://doi.org/10.1007/s40279-020-01282-z.
39. Guevara, S.A., Crunkhorn, M.L., Drew, M., Waddington, G., Périard, J.D., Etxebarria, N., Toohey, L.A., and Charlton, P. (2023). Injury and illness in short-course triathletes: a systematic review. J. Sport Health Sci. https://doi.org/10.1016/j.jshs.2023.03.002.
40. Liberati, A., Altman, D.G., Tetzlaff, J., Mulrow, C., Gøtzsche, P.C., Ioannidis, J.P.A., Clarke, M., Devereaux, P.J., Kleijnen, J., and Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. BMJ *339*, b2700. https://doi.org/10.1136/bmj.b2700.
41. Wang, H., and Chow, S.-C. (2007). Sample size calculation for comparing proportions. Wiley Encycl. Clin. Trials, 1–11. https://doi.org/10.1002/9780471462422.eoct005.
42. Cohen, J. (1988). In Statistical Power Analysis for the Behavioral Sciences, 2nd ed., J. Cohen, ed. (Lawrence Erlbaum Associates).
43. (2022). Two sample proportion test calculator. Stat. Kingdom *2*, 12–15. https://www.statskingdom.com/121proportion_normal2.html.
44. Ioannidis, J.P.A., Greenland, S., Hlatky, M.A., Khoury, M.J., Macleod, M.R., Moher, D., Schulz, K.F., and Tibshirani, R. (2014). Research: increasing value, reducing waste 2: Increasing value and reducing waste in research design, conduct, and analysis. Lancet *383*, 166–175. https://doi.org/10.1016/S0140-6736(13)62227-8.
45. Preobrazenski, N., Seigel, J., Halliday, S., Janssen, I., and McGlory, C. (2023). Single-leg disuse decreases skeletal muscle strength, size, and power in uninjured adults: A systematic review and meta-analysis. J. Cachexia Sarcopenia Muscle *14*, 684–696. https://doi.org/10.1002/jcsm.13201.
46. Chan, A.W., Song, F., Vickers, A., Jefferson, T., Dickersin, K., Gøtzsche, P.C., Krumholz, H.M., Ghersi, D., and Van Der Worp, H.B. (2014). Increasing value and reducing waste: Addressing inaccessible research. Lancet *383*, 257–266. https://doi.org/10.1016/S0140-6736(13)62296-5.
47. 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, et al.. (2020). Moving sport and exercise science forward: a call for the adoption of more transparent research practices. Sports Med. *50*, 449–459. https://doi.org/10.1007/s40279-019-01227-1.
48. Tiller, N.B., and Ekkekakis, P. (2023). Overcoming the "Ostrich Effect": a narrative review on the incentives and consequences of questionable research practices in kinesiology. Kinesiol. Rev. *12*, 201–216. https://doi.org/10.1123/kr.2022-0039.
49. Soderberg, C.K., Errington, T.M., Schiavone, S.R., Bottesini, J., Thorn, F.S., Vazire, S., Esterling, K.M., and Nosek, B.A. (2021). Initial evidence of research quality of registered reports compared with the standard publishing model. Nat. Hum. Behav. *5*, 990–997. https://doi.org/10.1038/s41562-021-01142-4.
50. Boutron, I., Altman, D.G., Moher, D., Schulz, K.F., Ravaud, P., CONSORT NPT Group, Golub, R.M., Grimshaw, J.M., Groves, T., Hopewell, S., et al. (2017). CONSORT Statement for randomized trials of nonpharmacologic treatments: A 2017 update and a CONSORT extension for nonpharmacologic trial abstracts. Ann. Intern. Med. *167*, 40–47. https://doi.org/10.7326/M17-0046.
51. Chan, A.W., Tetzlaff, J.M., Gøtzsche, P.C., Altman, D.G., Mann, H., Berlin, J.A., Dickersin, K., Hróbjartsson, A., Schulz, K.F., Parulekar, W.R., et al. (2013). SPIRIT 2013 explanation and elaboration: guidance for protocols of clinical trials. BMJ *346*, e7586. https://doi.org/10.1136/bmj.e7586.
52. Hoffmann, T.C., Glasziou, P.P., Boutron, I., Milne, R., Perera, R., Moher, D., Altman, D.G., Barbour, V., Macdonald, H., Johnston, M., et al. (2014). Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. BMJ *348*, g1687. https://doi.org/10.1136/bmj.g1687.
53. Slade, S.C., Dionne, C.E., Underwood, M., Buchbinder, R., Beck, B., Bennell, K., Brosseau, L., Costa, L., Cramp, F., Cup, E., et al. (2016). Consensus on Exercise Reporting Template (CERT): modified delphi study. Phys. Ther. *96*, 1514–1524. https://doi.org/10.2522/ptj.20150668.
54. Moher, D., Shamseer, L., Cobey, K.D., Lalu, M.M., Galipeau, J., Avey, M.T., Ahmadzai, N., Alabousi, M., Barbeau, P., Beck, A., et al. (2017). Stop this waste of people, animals and money. Nature *549*, 23–25. https://doi.org/10.1038/549023a.
55. Abt, G., Boreham, C., Davison, G., Jackson, R., Nevill, A., Wallace, E., and Williams, M. (2020). Power, precision, and sample size estimation in sport and exercise science research. J. Sports Sci. *38*, 1933–1935. https://doi.org/10.1080/02640414.2020.1776002.
56. Twomey, R., Yingling, V., Warne, J., Schneider, C., McCrum, C., Atkins, W., Murphy, J., Romero Medina, C., Harlley, S., and Caldwell, A. (2021). The nature of our literature: a registered report on the positive result rate and reporting practices in kinesiology. Commun. Kinesiol. *1*. https://doi.org/10.51224/cik.v1i3.43.
57. Freccia, G.W., Carvalho, T.D., Catarina, S., and Simone, R.P. (2022). A reproducibility or an identity crisis in sports and exercise sciences? A viewpoint on current challenges and solutions. Preprint at SportRxiv. https://doi.org/10.51224/SRXIV.171.
58. Sainani, K., and Chamari, K. (2022). Wish list for improving the quality of statistics in sport science. Int. J. Sports Physiol. Perform. *17*, 673–674. https://doi.org/10.1123/ijspp.2022-0023.
59. Mesquida, C., Murphy, J., Lakens, D., and Warne, J. (2022). Replication concerns in sports and exercise science: a narrative review of selected methodological issues in the field. R. Soc. Open Sci. *9*, 220946. https://doi.org/10.1098/rsos.220946.
60. Borg, D.N., Bon, J.J., Sainani, K.L., Baguley, B.J., Tierney, N.J., and Drovandi, C. (2020). Comment on: 'Moving sport and exercise science forward: a call for the adoption of more transparent research practices. Sports Med. *50*, 1551–1553. https://doi.org/10.1007/s40279-020-01298-5.
61. Bala, M.M., Akl, E.A., Sun, X., Bassler, D., Mertz, D., Mejza, F., Vandvik, P.O., Malaga, G., Johnston, B.C., Dahm, P., et al. (2013). Randomized trials published in higher vs. lower impact journals differ in design, conduct, and analysis. J. Clin. Epidemiol. *66*, 286–295. https://doi.org/10.1016/j.jclinepi.2012.10.005.
62. Berger, V.W., and Exner, D.V. (1999). Detecting selection bias in randomized clinical trials. Control. Clin. Trials *20*, 319–327. https://doi.org/10.1016/S0197-2456(99)00014-8.
63. Altman, D.G., and Bland, J.M. (1999). Statistics notes: treatment allocation in controlled trials: why randomise? BMJ *318*, 1209. https://doi.org/10.1136/bmj.318.7192.1209.
64. Altman, D.G., and Schulz, K.F. (2001). Statistics notes: Concealing treatment allocation in randomised trials. Br. Med. J. *323*, 446–447. https://doi.org/10.1136/bmj.323.7310.446.
65. Berger, V.W., and Christophi, C.A. (2003). Randomization technique, allocation concealment, masking, and susceptibility of trials to selection bias. J. Mod. Appl. Stat. Methods *2*, 80–86. https://doi.org/10.22237/jmasm/1051747680.
66. Hecksteden, A., Faude, O., Meyer, T., and Donath, L. (2018). How to Construct, Conduct and Analyze an Exercise Training Study?. https://doi.org/10.3389/fphys.2018.01007.
67. Boutron, I., Tubach, F., Giraudeau, B., and Ravaud, P. (2004). Blinding was judged more difficult to achieve and maintain in nonpharmacologic than pharmacologic trials. J. Clin. Epidemiol. *57*, 543–550. https://doi.org/10.1016/j.jclinepi.2003.12.010.
68. Schulz, K.F., Chalmers, I., and Altman, D.G. (2002). The landscape and lexicon of blinding in randomized trials. Ann. Intern. Med. *136*, 254–259. https://doi.org/10.7326/0003-4819-136-3-200202050-00022.
69. Preobrazenski, N., Islam, H., Drouin, P.J., Bonafiglia, J.T., Tschakovsky, M.E., and Gurd, B.J. (2020). A novel gravity-induced blood flow restriction model augments ACC phosphorylation and PGC-1α mRNA in human skeletal muscle following aerobic exercise: a randomized crossover study. Appl. Physiol. Nutr. Metab. *45*, 641–649. https://doi.org/10.1139/apnm-2019-0641.

70. Nunan, D., Aronson, J., and Bankhead, C. (2018). Catalogue of bias: attrition bias. BMJ Evid. Based. Med. *23*, 21–22. https://doi.org/10.1136/ebmed-2017-110883.

71. Dwan, K., Gamble, C., Williamson, P.R., and Kirkham, J.J.; Reporting Bias Group (2013). Systematic review of the empirical evidence of study publication bias and outcome reporting bias — an updated review. PLoS One *8*, e66844. https://doi.org/10.1371/journal.pone.0066844.

72. Dickersin, K., and Rennie, D. (2003). Registering clinical trials. JAMA *290*, 516–523. https://doi.org/10.1001/jama.290.4.516.

73. Vankov, I., Bowers, J., and Munafò, M.R. (2014). On the persistence of low power in psychological science. Q. J. Exp. Psychol. *67*, 1037–1040. https://doi.org/10.1080/17470218.2014.885986.

74. Ekkekakis, P., Swinton, P., and Tiller, N.B. (2023). Extraordinary Claims in the literature on high - intensity interval training (HIIT): I . bonafide scientific revolution or a looming crisis of replication and credibility? Sports Med. *53*, 1865–1890. https://doi.org/10.1007/s40279-023-01880-7.

75. Altman, D.G., and Bland, J.M. (1995). Absence of evidence is not evidence of absence. BMJ *311*, 485. https://doi.org/10.1136/bmj.311.7003.485.

76. Walter, S.D., and Hart, L.E. (1990). Application of epidemiological methodology to sports and exercise science research. Exerc. Sport Sci. Rev. *18*, 417–448.

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

78. Minozzi, S., Cinquini, M., Gianola, S., Gonzalez-Lorenzo, M., and Banzi, R. (2020). The revised Cochrane risk of bias tool for randomized trials (RoB 2) showed low interrater reliability and challenges in its application. J. Clin. Epidemiol. *126*, 37–44. https://doi.org/10.1016/j.jclinepi.2020.06.015.

79. Sterne, J.A.C., Savović, J., Page, M.J., Elbers, R.G., Blencowe, N.S., Boutron, I., Cates, C.J., Cheng, H.-Y., Corbett, M.S., Eldridge, S.M., et al. (2019). RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ *366*, l4898. https://doi.org/10.1136/bmj.l4898.

80. Soares, H.P., Daniels, S., Kumar, A., Clarke, M., Scott, C., Swann, S., and Djulbegovic, B.; Radiation Therapy Oncology Group (2004). Bad reporting does not mean bad methods for randomised trials: observational study of randomised controlled trials performed by the Radiation Therapy Oncology Group. Br. Med. J. *328*, 22–24. https://doi.org/10.1136/bmj.328.7430.22.

81. Mhaskar, R., Djulbegovic, B., Magazin, A., Soares, H.P., and Kumar, A. (2012). Published methodological quality of randomized controlled trials does not reflect the actual quality assessed in protocols. J. Clin. Epidemiol. *65*, 602–609. https://doi.org/10.1016/j.jclinepi.2011.10.016.

82. Kolaski, K., Logan, L.R., and Ioannidis, J.P.A. (2023). Guidance to best tools and practices for systematic reviews. BMC Infect. Dis. *23*, 383–429. https://doi.org/10.1186/s12879-023-08304-x.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
Microsoft Excel	Microsoft	

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Dr. Brendon Gurd (gurdb@queensu.ca).

Material availability

This study did not generate new unique reagents.

Data and code availability

All data used to perform analyses have been included as publicly available supplemental information. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request. This paper does not report original code.

METHOD DETAILS

The current study is a systematic review, and it is not pre-clinical research. The intended methods were documented on Open Science Framework (osf.io/jznv8) on April 29, 2021 prior to the literature searches, data extraction, and data analysis. The registration was updated on July 8, 2022 to include a finalized analysis plan. The NIH and ARRIVE reporting guidelines for preclinical work do not apply. All methods are detailed in our manuscript and supplemental files.

QUANTIFICATION AND STATISTICAL ANALYSIS

Descriptive statistics were used for domain, type of exercise, type of training, and categories of bias in Microsoft Excel (v.16.72). For selection bias, we used the higher number of unclear/high judgements from its constituents (sequence generation and allocation concealment). Although our initial intention was to assess risk of bias in all included studies, we modified our assessment approach after the full-text review to assess 170 studies published in 1995 and 170 studies published in 2020 (modification was included in an update to our study registration on July 8, 2022). This sample size was selected because a sample size of 167 provided 80% power to detect a 15% difference in proportions between years with a 95% confidence level using the formula from section 3.1 in Wang and Chow.[41] We used lower-than-expected proportions of unclear/high studies (e.g., 65% vs 50%; Cohen's h=0.3) in our power calculation because this yielded a more conservative sample size estimate compared with using a 15% difference at higher population proportions (see Table 6.2.1 in Cohen[42]).

Two-tailed z-score tests for two population proportions with no continuity correction[43] were performed to determine whether the proportions of studies judged to have unclear or high risks of bias were different between studies published in 1995 and 2020. The Bonferroni correction was applied to account for multiple comparisons and avoid spurious positives (Bonferroni corrected α=0.05/5 types of bias=0.01). Significance was therefore set at p<0.01.