Meta-science in physiotherapy research: a structured review of what we know and what we still do not

Jürgen Degenfellner

Institute of Physiotherapy, ZHAW Zurich University of Applied Sciences, Winterthur, Switzerland

degn@zhaw.ch · ORCID 0000-0001-6886-4583

Abstract

Introduction. Physiotherapy has built a large clinical evidence base, but whether that evidence can be trusted depends not only on the trials inside it but on how the research is conducted, reported, and shared. Meta-research, the empirical study of research itself, makes those questions answerable, yet its findings for physiotherapy have never been drawn together.

Methods. I carried out a structured review that maps physiotherapy-specific meta-research across the recognised themes of the field (methods, reporting, reproducibility, evaluation, and incentives). I searched PubMed, Scopus, and Google Scholar in June 2026 with domain-specific terms, charted the evidence by domain, and rated each domain's maturity. Where no physiotherapy-specific evidence existed, landmark general biomedical work is cited and flagged as borrowed.

Results. The picture is uneven. Trial quality has risen slowly; nine in ten musculoskeletal reviews are critically low on AMSTAR 2; spin appears in close to nine in ten "non-significant" trial reports and shifts clinicians' interpretation; baseline characteristics are tested for "significance" in three-fifths of recent trials; only one-third of trials are usefully registered; and replication is essentially absent while open data are rare. Conflicts of interest, citation bias, peer-review quality, the incentive and evaluation systems, and the use of large language models remain largely or wholly unstudied. No field-wide synthesis appears to exist.

Conclusions. Physiotherapy meta-research is real but young and patchy. I map what is known,

separate genuine evidence gaps from unstudied ground, and propose a concrete agenda for strengthening research integrity, transparency, and reproducibility.

Keywords: Bias; Physical Therapy Modalities; Reproducibility of Results; Research Design; Review Literature as Topic; Statistics as Topic.

What is already known about this topic. Single audits have flagged problems in physiotherapy research, including low methodological quality, frequent spin, and poor data sharing, but these have never been synthesised or placed within the wider field of meta-research.

What does this review add. It provides a first field-wide map of meta-research in physiotherapy, distinguishes genuine evidence gaps from unstudied territory, and sets out a concrete agenda for improving research integrity, transparency, and reproducibility.

1 Introduction

Evidence-based practice arrived in physiotherapy with real ambition. The PEDro database, launched in 1999, now indexes tens of thousands of controlled trials; systematic reviews appear faster than any clinician can read them; and guidelines for almost every musculoskeletal complaint now rest on pooled trial data. The promise was that better evidence would mean better care.

There is a catch, and it is one that medicine as a whole has been forced to confront. A body of evidence is only as good as the studies inside it, and a growing literature suggests that many studies, across many fields, are less dependable than their confident abstracts imply. John Ioannidis put the provocation most sharply in 2005 with the claim that most published research findings are false [1], though Douglas Altman had warned a decade earlier of a quieter scandal, that of medical research done badly rather than dishonestly [2]. Iain Chalmers and Paul Glasziou later estimated that around 85 per cent of the worldwide investment in biomedical research is avoidably wasted, lost to poor questions, poor design, poor reporting, and results that are never made usable [3–5]. When psychologists tried to repeat one hundred of their own headline findings, fewer than half came out the same way [6]; cancer biologists fared little better [7].

These problems have given rise to a discipline of their own. *Meta-research*, or meta-science, is research on research: the empirical study of how science is performed, reported, reproduced, evaluated, and rewarded. Ioannidis and colleagues sort it into those five themes [8–10]. It treats the scientific literature itself as data. Rather than asking whether a particular exercise programme reduces back pain, meta-research asks whether the trials of that programme were registered before they began, whether their results were reported in full, whether their statistics were sound, and whether anyone could reproduce them.

Physiotherapy is an unusually good place to ask such questions, and an unusually neglected one. Its interventions are complex and hard to blind; its outcomes are often subjective; its trials are frequently small; and much of its literature sits behind paywalls in specialist journals that receive little outside scrutiny. These features make physiotherapy more exposed to the problems meta-research studies, not less. Yet while general medicine, psychology, and preclinical biology have all been examined at length, the physiotherapy literature has been audited only in scattered patches. This review draws those patches together. It asks whether a field-wide overview already exists, maps the domains where physiotherapy-specific evidence has accumulated, names plainly the domains where it has not, and moves from limitations to proposed solutions.

2 Methods

This is a structured review aimed at mapping a broad field and locating its gaps, in the tradition of scoping reviews rather than focused systematic reviews. Its objective was to chart what meta-research exists for physiotherapy across the five themes above and to identify where physiotherapy-specific evidence is absent. As in a scoping review, I did not appraise the risk of bias of individual studies or pool effects. I also state its pragmatic limits plainly: the review was not pre-registered, the search was structured but not exhaustive, and I screened and charted the evidence alone rather than in duplicate. I drew on the principles of PRISMA-ScR and on the SANRA guidance where they applied. Interpretation is therefore part of the method, with the familiar risk that an author finds what an author expects, and the reader should weigh the argument accordingly.

The literature was assembled through structured searches of PubMed, Scopus, and Google Scholar, supplemented by citation tracking and a set of automated, multi-source searches across the domains below. Priority went to physiotherapy-specific studies, meaning audits of

physiotherapy, physical therapy, or rehabilitation journals and trials. Where no physiotherapy-specific evidence existed, the relevant landmark work from general biomedical meta-research is cited and flagged as borrowed rather than native. One of the most data-rich sources is a preprint at the time of writing and is labelled as such. Where this review reports that a topic is unstudied in physiotherapy, that judgement rests on a dedicated search carried out in June 2026 using terms specific to each domain; absence means that no qualifying physiotherapy-specific study was found, not that none can exist, and I name the closest adjacent work wherever it helps the reader judge.

Field-level meta-research overviews now exist for neighbouring disciplines, including a recent scoping review of methodological issues in rehabilitation [11] and an overview of meta-research in neuroscience [12]. I found no comparable field-wide review of meta-research in physiotherapy. The evidence below is organised loosely around the five themes [8]; the map is fullest for methods, reporting, and reproducibility, and thinner for evaluation and incentives, where I have tried to be explicit that the gap is partly in my own coverage and not only in the literature. Meta-research is itself a moving field, and no single review will capture every corner of it.

3 What we already know

The mature strands of physiotherapy meta-research cluster around a handful of questions about trustworthiness: how good the studies are, how faithfully they are reported, how their results are framed, how sound their statistics are, and how openly they are conducted. Each has produced findings worth taking seriously.

3.1 Methodological quality and risk of bias

The longest-running evidence concerns trial quality. Using the PEDro scale, Moseley et al. [13] examined 10,025 physiotherapy trials published between 1960 and 2009 and found that methodological quality had improved, but at the unhurried pace of about 0.6 points per decade on the ten-point PEDro scale. At that rate the average trial would take the better part of a century to become convincingly rigorous. Systematic reviews fare no better, which matters because they are produced in ever greater numbers and increasingly drive practice [14]. Applying the AMSTAR 2 appraisal tool [15] to one hundred randomly selected reviews of musculoskeletal

physiotherapy interventions, Ferri et al. [16] rated 90 per cent as critically low in confidence, with no sign of improvement across the decade studied. The instruments are themselves not interchangeable: comparing the Cochrane risk-of-bias tool with the PEDro scale across 1,442 trials from 108 Cochrane reviews, Moseley et al. [17] found only modest item agreement and poor agreement on summary scores, a reminder that a trial's apparent quality depends partly on the ruler used to measure it.

3.2 Reporting completeness and guideline adherence

A trial that cannot be understood cannot be replicated. Reporting guidelines such as CONSORT for trials [18], PRISMA for systematic reviews [19], and the TIDieR checklist for interventions [20] were written to enforce that clarity, and they have had limited effect in physiotherapy. Comparing intervention descriptions from 2000 with those from 2018, McCambridge et al. [21] found that completeness on the TIDieR checklist rose only from 7.5 to 8.3 out of fourteen, an improvement of around five per cent and well short of the 20 per cent difference the authors had set in advance as meaningful. Two decades and a dedicated checklist later, a reader still often cannot tell what the patients received. The wider record gives little ground for optimism: across nearly 21,000 trials in Cochrane reviews, poor reporting and inadequate methods improved only slowly over time [22].

A related question is not how completely outcomes are reported but which outcomes are chosen. Here physiotherapy has been constructive: consensus core outcome sets now exist for major conditions such as non-specific low back pain, specifying both the domains worth measuring and the instruments to measure them [23, 24]. They are a direct answer to the outcome heterogeneity that makes trials hard to combine, though how widely new trials adopt them has yet to be audited.

3.3 Spin and the framing of results

If the numbers in a trial are fixed, the words around them are not. *Spin*, the reporting of a study so as to make its results look more favourable than the data warrant, was named and defined for randomised trials by Boutron et al. [25], who showed it arises most often when a "non-significant" primary outcome is quietly talked up. Physiotherapy is no exception. Auditing 44 physiotherapy trials with "non-significant" primary outcomes, Kumar et al. [26] found spin in

90 per cent of abstracts and 88 per cent of main-text conclusions. What lifts this from irritation to genuine concern is that spin works: in a blinded randomised experiment, Khanpara and Prakash [27] asked physiotherapists to rate, on a nought-to-ten scale, how much a treatment had helped, judging it from the abstract of a trial whose primary outcome was "non-significant". The two groups saw the same abstract, differing only in its spin, and those given the spun version rated the benefit 1.8 points higher on average (6.1 versus 4.3; 95% CI 0.8 to 2.8), a medium-sized effect. The framing alone shifted how well clinicians judged the treatment to have worked.

3.4 Statistical fragility and the significance habit

Physiotherapy trials are often small, and small trials make fragile claims, a lesson neuroscience learned the hard way when its underpowered studies proved unreliable [28]. Examining high-quality sports physiotherapy trials from the PEDro database, Bleakley et al. [29] estimated the false-positive risk behind their "significant" findings: the median risk was about 9 per cent, but 63 per cent of "significant" results carried a risk above the conventional 5 per cent threshold, and almost one in five carried a risk above 50 per cent, meaning the claimed effect was as likely as not to be noise. These were the better trials, chosen for their high methodological scores. The habit underneath is the reflexive reliance on statistical significance, what Gerd Gigerenzer called the "mindless" ritual of running the test without thinking about what it answers [30], and one the wider statistical community has spent a decade trying to break [31–33]. A 2025 preprint audited 465 articles in the national-association journals of the United States, Canada, the United Kingdom, and Australia and found the habit alive and well: 63 per cent of quantitative studies leaned primarily on significance testing, only a third stated a hypothesis at all, hypotheses were reported as supported in 83 per cent of cases, and 39 per cent of "significant" p-values fell in the fragile band just below 0.05, against the 26 per cent one would expect from adequately powered studies [34]. Patterns like these are the signature of questionable research practices; the sister field of sport and exercise medicine put a number on it, with hypotheses supported in 82 per cent of papers, a rate too high to be honest [35].

3.5 Registration, outcome switching, and selective reporting

Prospective registration, and the protocols good practice now expects alongside it [36], were introduced so that a trial's intended outcomes could be compared with what it eventually reported, closing off the temptation to promote whichever measure reached "significance". In physiotherapy the mechanism is only half built. Across 138 musculoskeletal trials in fifteen specialist journals, Riley et al. [37] found that only one-third were prospectively registered with stated intentions consistent with the published report; for most of the rest, consistency could not be checked. Registration has grown more common over time, but prospective, usable registration remains the exception [38, 39].

3.6 Open science: data, code, and replication

The clearest single snapshot of physiotherapy's open-science culture comes from the same 2025 preprint [34]: of 465 recent articles in four flagship journals, 28 per cent were preregistered, 12 per cent shared their data, and, tellingly, none were replication studies or registered reports. A field that never repeats its own experiments has no internal check on whether its findings hold, and the open-research practices designed to make such checks possible [40] are barely in evidence. A parallel audit of 103 physical-activity systematic reviews told the same story from the synthesis side, with fewer than five per cent reporting a public protocol [41]. How far this extends to the many specialist journals current audits cannot reach, and to code as well as data, remains open.

3.7 Research waste, retraction, and predation

The waste Chalmers and Glasziou described has a quiet local form. A new study is justified only if it addresses a genuine gap, which is why leading journals increasingly ask authors to cite a systematic review showing the question is still open; Rosen et al. [42] found that not one of 149 physiotherapy journals required this in their author guidelines. At the far edge sit retraction and predation. A bibliometric analysis identified 280 retracted articles in rehabilitation between 1984 and 2020, a number the authors suspected was only the visible tip [43], and trials in predatory physiotherapy journals scored markedly lower on the PEDro scale than those in legitimate open-access journals [44], confirming that where peer review is weakest, the research

Table 1: The state of meta-research in physiotherapy, by domain.

Domain	Representative physiotherapy evidence	Maturity
Trial methodological quality	PEDro quality rising ∼0.6 pts/decade [13]	Established
Systematic review quality	90% critically low on AMSTAR 2 [16]	Established
Reporting completeness	TIDieR barely improved 2000–2018 [21]	Established
Core outcome sets	Consensus sets developed (e.g. low back pain) [23]	Established
Spin and result framing	Spin in ∼88–90% of trials; alters clinician reading [26, 27]	Established
Statistical fragility	63% of "significant" findings exceed 5% false-positive risk [29]	Emerging
Registration / outcome switching	Only ∼1/3 of trials usefully prospectively registered [37]	Emerging
Open data, code, replication	12% open data, 0% replications [34]	Emerging
Retraction / predatory journals	280 rehabilitation retractions; lower quality in predatory journals [43, 44]	Emerging
Baseline significance testing	61% of trials by 2018 [45]	Emerging
Overlapping-CI / normality misuse	Uncounted; commentary only [46]	Absent
Bayesian methods uptake	Prevalence never estimated in physiotherapy	Absent
Conflicts of interest / funding	One sponsorship-bias study; disclosure unaudited [47]	Sparse
Authorship integrity	Honorary authorship and inflation measured [48, 49]	Sparse
Citation distortion	Reference accuracy only, not bias [50]	Absent
Peer-review quality	Essentially unstudied	Absent
Research evaluation & incentives	Citation-determinant bibliometrics only [51]	Absent
AI and large language models	No empirical study; one perspective [52]	Absent

is weakest too.

4 What is still missing

The domains above are the lit portion of the map. The dark portion is larger, and naming it precisely matters more than gesturing at it (Table 1). Several gaps stand out.

Specific statistical errors are only half-counted. The *Table 1 fallacy*, testing baseline characteristics for "significance" in a randomised trial, was shown to be logically incoherent decades ago [53, 54] and is discouraged by CONSORT, yet it remains common; a 2025 audit found it in nearly half of anaesthesiology trials [55], and physiotherapy is, if anything, worse, with baseline "significance" testing in 61 per cent of trials by 2018, up from a third in 2000

[45]. What remains uncounted is how often a spurious baseline "imbalance" then changes the analysis. Two related errors have escaped measurement altogether: the *overlapping confidence interval fallacy* [56], which physiotherapy has aired in a commentary [46] but never counted, and the misuse of normality tests, which has no prevalence estimate at all. The toolkit is also narrow: physiotherapy appears almost entirely frequentist, but even that is an impression, since no estimate exists of how often Bayesian or other methods appear.

Incentives and integrity are barely examined. Here the evidence is thin rather than absent. Industry funding has been looked at once, in a study of 393 rehabilitation trials that found a weak signal of larger effects where the sponsor's influence was unclear [47], but no audit has asked how often physiotherapy papers disclose conflicts of interest at all. Authorship has fared a little better, with a study of honorary authorship in rehabilitation journals, which a strict ICMJE definition put at 55 per cent of surveyed authorships [48], and one of authorship inflation, measured through letters to the editor, in exercise-science and physiotherapy journals [49]. But citation behaviour has been examined only for reference accuracy, not bias [50]; the quality and consistency of peer review in physiotherapy journals appears entirely unstudied; and beyond a bibliometric look at what makes a trial get cited [51], no one has examined how the field's outputs are evaluated, how the impact factor is misused in hiring and promotion, or how the publish-or-perish economy drives salami-slicing and authorship pressure. These reward and evaluation systems are much of the reason the other problems persist.

The arrival of machine writing is unmeasured. Large language models are already used to draft, translate, and review scientific text across biomedicine, and automated summaries have been shown to overstate what studies found. How far this has reached physiotherapy has not been studied; the only published engagement is a perspective urging clinicians to adopt the tools [52], not an audit of how they are reshaping the literature. Reliable detection of machine-written text remains an unsolved problem, which makes the question harder to answer, not easier to ignore.

Publication bias is asserted more than measured. The fragile-p-value signal hints strongly at selective publication [34], and the registration audits touch the same nerve from the side of outcome reporting [37, 39]. What the field lacks is direct measurement: funnel-

plot or small-study-effect analyses across physiotherapy meta-analyses, and tracking of how many registered trials reach print with their results intact. The suspicion is well founded; the measurement is thin.

5 Towards a research and reform agenda

The gaps suggest a programme that is as much editorial and institutional as academic.

1. **Make prospective registration and registered reports the norm.** Editors can require prospective registration as a condition of review and offer registered reports, which evaluate the question and methods before results exist and so blunt spin, outcome switching, and publication bias at the source.
2. **Enforce reporting guidelines rather than merely endorse them.** Physiotherapy journals can move from listing CONSORT, PRISMA, and TIDieR to checking them at submission, and meta-research can test whether enforcement changes what is published.
3. **Census the statistical fallacies.** Estimate, across a defined trial corpus, the prevalence of baseline "significance" testing, overlapping-confidence-interval misinterpretation, inappropriate normality testing, and underpowered designs, and treat them together as indicators of statistical culture.
4. **Measure openness routinely.** Extend automated data-and-code detection across the full journal landscape, including the paywalled specialist journals current audits cannot reach, and report trends over time.
5. **Open the integrity and incentive questions.** Build on the few existing studies of sponsorship and authorship to measure conflict-of-interest disclosure, peer-review quality, citation bias, and the metric-driven incentives that shape behaviour.
6. **Take artificial intelligence seriously.** Study how language models are reshaping the writing, reviewing, and reading of physiotherapy research, and what disclosure norms and editorial safeguards the field needs, before the practice becomes universal.
7. **Build shared infrastructure.** An open corpus of physiotherapy trials and reviews, with reusable extraction tools, would let any of the above be done once and repeated cheaply, turning one-off audits into continuous monitoring.

6 Conclusions

It is tempting to read a review like this as an indictment, and it is not meant as one. Every problem named here was named first in general medicine, often by the very people who built evidence-based practice. A discipline mature enough to study its own weaknesses is in a stronger position than one that assumes it has none.

The honest summary is that physiotherapy meta-research is real but young, strong in a few places and silent in many. Its firmest findings, on reporting, spin, and openness, point the same way: the literature is less dependable than its confidence suggests, and improvement has been slow. Its largest gaps, on incentives, integrity, and the new machinery of automated text, are barely mapped. A field that knows precisely where it is weak can choose where to look next, and can act, through the editorial and institutional levers above, to raise the standard of the evidence that patients ultimately depend on.

References

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

[2] Douglas G. Altman. The scandal of poor medical research. *BMJ*, 308(6924):283–284, 1994. doi: 10.1136/bmj.308.6924.283.

[3] Iain Chalmers and Paul Glasziou. Avoidable waste in the production and reporting of research evidence. *The Lancet*, 374(9683):86–89, 2009. doi: 10.1016/S0140-6736(09)60329-9.

[4] Paul Glasziou, Douglas G. Altman, Patrick Bossuyt, Isabelle Boutron, Mike Clarke, Steven Julious, Susan Michie, David Moher, and Elizabeth Wager. Reducing waste from incomplete or unusable reports of biomedical research. *The Lancet*, 383(9913):267–276, 2014. doi: 10.1016/S0140-6736(13)62228-X.

[5] Malcolm R. Macleod, Susan Michie, Ian Roberts, Ulrich Dirnagl, Iain Chalmers, John P. A. Ioannidis, Rustam Al-Shahi Salman, An-Wen Chan, and Paul Glasziou. Biomedical research: Increasing value, reducing waste. *The Lancet*, 383(9912):101–104, 2014. doi: 10.1016/S0140-6736(13)62329-6.

[6] Open Science Collaboration. Estimating the reproducibility of psychological science. *Science*, 349(6251):aac4716, 2015. doi: 10.1126/science.aac4716.

[7] C. Glenn Begley and Lee M. Ellis. Raise standards for preclinical cancer research. *Nature*, 483(7391):531–533, 2012. doi: 10.1038/483531a.

[8] John P. A. Ioannidis, Daniele Fanelli, Debbie Drake Dunne, and Steven N. Goodman. Meta-research: Evaluation and improvement of research methods and practices. *PLoS Biology*, 13(10):e1002264, 2015. doi: 10.1371/journal.pbio.1002264.

[9] Tom E. Hardwicke, Stylianos Serghiou, Perrine Janiaud, Valentin Danchev, Sophia Crüwell, Steven N. Goodman, and John P. A. Ioannidis. Calibrating the scientific ecosystem through meta-research. *Annual Review of Statistics and Its Application*, 7:11–37, 2020. doi: 10.1146/annurev-statistics-031219-041104.

[10] Marcus R. Munafò, Brian A. Nosek, Dorothy V. M. Bishop, Katherine S. Button, Christopher D. Chambers, Nathalie Percie du Sert, Uri Simonsohn, Eric-Jan Wagenmakers, Jennifer J. Ware, and John P. A. Ioannidis. A manifesto for reproducible science. *Nature Human Behaviour*, 1:0021, 2017. doi: 10.1038/s41562-016-0021.

[11] Chiara Arienti, Susan Armijo-Olivo, Silvia Minozzi, Lisa Tjosvold, Stefano Giuseppe Lazzarini, Michele Patrini, and Stefano Negrini. Methodological issues in rehabilitation research: A scoping review. *Archives of Physical Medicine and Rehabilitation*, 102(8): 1614–1622, 2021. doi: 10.1016/j.apmr.2021.04.006.

[12] Ivan David Lozada-Martinez, Luis Miguel Delgado Arias, Diana Marcela Perea Rojas, and Indiana Luz Rojas Torres. Meta-research in neuroscience: An urgent call to strengthen the reliability and translation of knowledge into evidence-based neurological practice. *Journal of Clinical Medicine*, 14(23):8552, 2025. doi: 10.3390/jcm14238552.

[13] Anne M. Moseley, Robert D. Herbert, Christopher G. Maher, Catherine Sherrington, and Mark R. Elkins. Reported quality of randomized controlled trials of physiotherapy interventions has improved over time. *Journal of Clinical Epidemiology*, 64(6):594–601, 2011. doi: 10.1016/j.jclinepi.2010.08.009.

[14] John P. A. Ioannidis. The mass production of redundant, misleading, and conflicted

systematic reviews and meta-analyses. *The Milbank Quarterly*, 94(3):485–514, 2016. doi: 10.1111/1468-0009.12210.

[15] Beverley J. Shea, Barnaby C. Reeves, George Wells, Micere Thuku, Candyce Hamel, Julian Moran, David Moher, Peter Tugwell, Vivian Welch, Elizabeth Kristjansson, and David A. Henry. AMSTAR 2: A critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ*, 358:j4008, 2017. doi: 10.1136/bmj.j4008.

[16] Nicola Ferri, Elisa Ravizzotti, Alessandro Bracci, Giulia Carreras, Paolo Pillastrini, and Mauro Di Bari. The confidence in the results of physiotherapy systematic reviews in the musculoskeletal field is not increasing over time: a meta-epidemiological study using amstar 2 tool. *Journal of Clinical Epidemiology*, 169:111303, 2024. doi: 10.1016/j.jclinepi.2024.111303.

[17] Anne M. Moseley, Prinon Rahman, George A. Wells, Joshua R. Zadro, Catherine Sherrington, Karine Toupin-April, and Lucie Brosseau. Agreement between the cochrane risk of bias tool and physiotherapy evidence database (pedro) scale: A meta-epidemiological study of randomized controlled trials of physical therapy interventions. *PLOS ONE*, 14(9): e0222770, 2019. doi: 10.1371/journal.pone.0222770.

[18] Kenneth F. Schulz, Douglas G. Altman, and David Moher. CONSORT 2010 statement: Updated guidelines for reporting parallel group randomised trials. *BMJ*, 340:c332, 2010. doi: 10.1136/bmj.c332.

[19] Matthew J. Page, Joanne E. McKenzie, Patrick M. Bossuyt, Isabelle Boutron, Tammy C. Hoffmann, Cynthia D. Mulrow, et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372:n71, 2021. doi: 10.1136/bmj.n71.

[20] Tammy C. Hoffmann, Paul P. Glasziou, Isabelle Boutron, Ruairidh Milne, Rafael Perera, David Moher, Douglas G. Altman, et al. Better reporting of interventions: Template for intervention description and replication (TIDieR) checklist and guide. *BMJ*, 348:g1687, 2014. doi: 10.1136/bmj.g1687.

[21] Alana B. McCambridge, Anthony M. Nasser, Poonam Mehta, Peter W. Stubbs, and Arianne P. Verhagen. Has reporting on physical therapy interventions improved in 2

decades? an analysis of 140 trials reporting on 225 interventions. *Journal of Orthopaedic & Sports Physical Therapy*, 51(10):503–509, 2021. doi: 10.2519/jospt.2021.10642.

[22] Agnes Dechartres, Ludovic Trinquart, Ignacio Atal, David Moher, Kay Dickersin, Isabelle Boutron, Elodie Perrodeau, Douglas G. Altman, and Philippe Ravaud. Evolution of poor reporting and inadequate methods over time in 20,920 randomised controlled trials included in cochrane reviews. *BMJ*, 357:j2490, 2017. doi: 10.1136/bmj.j2490.

[23] Alessandro Chiarotto, Richard A. Deyo, Caroline B. Terwee, Maarten Boers, Rachelle Buchbinder, Terry P. Corbin, Leonardo O. P. Costa, Nadine E. Foster, Margreth Grotle, Bart W. Koes, Francisco M. Kovacs, Chung-Wei Christine Lin, Christopher G. Maher, Adam M. Pearson, Wilco C. Peul, Mark L. Schoene, Dennis C. Turk, Maurits W. van Tulder, and Raymond W. Ostelo. Core outcome domains for clinical trials in non-specific low back pain. *European Spine Journal*, 24(6):1127–1142, 2015. doi: 10.1007/s00586-015-3892-3.

[24] Alessandro Chiarotto, Maarten Boers, Richard A. Deyo, Rachelle Buchbinder, Terry P. Corbin, Leonardo O. P. Costa, Nadine E. Foster, Margreth Grotle, Bart W. Koes, Francisco M. Kovacs, Chung-Wei Christine Lin, Christopher G. Maher, Adam M. Pearson, Wilco C. Peul, Mark L. Schoene, Dennis C. Turk, Maurits W. van Tulder, Caroline B. Terwee, and Raymond W. Ostelo. Core outcome measurement instruments for clinical trials in nonspecific low back pain. *Pain*, 159(3):481–495, 2018. doi: 10.1097/j.pain.0000000000001117.

[25] Isabelle Boutron, Susan Dutton, Philippe Ravaud, and Douglas G. Altman. Reporting and interpretation of randomized controlled trials with statistically nonsignificant results for primary outcomes. *JAMA*, 303(20):2058–2064, 2010. doi: 10.1001/jama.2010.651.

[26] J. Saravan Kumar, V. Prakash, K. Hariohm, and I. Naveen Kumar. Misreporting study results in main texts and abstracts of randomized controlled trials published by physical therapy researchers (spin). *Physiotherapy Practice and Research*, 44(2):109–115, 2023. doi: 10.3233/PPR-230747.

[27] Heppy Khanpara and V. Prakash. Effect of spin in the abstract of a randomised controlled trial on physiotherapists' perception of treatment benefit: a randomised controlled trial. *BMJ Evidence-Based Medicine*, 27(2):97–103, 2022. doi: 10.1136/bmjebm-2021-111714.

[28] Katherine S. Button, John P. A. Ioannidis, Claire Mokrysz, Brian A. Nosek, Jonathan Flint, Emma S. J. Robinson, and Marcus R. Munafò. Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, 14(5):365–376, 2013. doi: 10.1038/nrn3475.

[29] Chris Bleakley, Jonathan Reijgers, and James M. Smoliga. Many high-quality randomized controlled trials in sports physical therapy are making false-positive claims of treatment effect: A systematic survey. *Journal of Orthopaedic & Sports Physical Therapy*, 50(2): 104–109, 2020. doi: 10.2519/jospt.2020.9264.

[30] Gerd Gigerenzer. Mindless statistics. *The Journal of Socio-Economics*, 33(5):587–606, 2004. doi: 10.1016/j.socec.2004.09.033.

[31] Ronald L. Wasserstein and Nicole A. Lazar. The ASA statement on p-values: Context, process, and purpose. *The American Statistician*, 70(2):129–133, 2016. doi: 10.1080/00031305.2016.1154108.

[32] Ronald L. Wasserstein, Allen L. Schirm, and Nicole A. Lazar. Moving to a world beyond "p $<$ 0.05". *The American Statistician*, 73(sup1):1–19, 2019. doi: 10.1080/00031305.2019.1583913.

[33] Valentin Amrhein, Sander Greenland, and Blake McShane. Scientists rise up against statistical significance. *Nature*, 567(7748):305–307, 2019. doi: 10.1038/d41586-019-00857-9.

[34] François Jabouille et al. Replicability and transparency in physical therapy research: Time to wake up. *medRxiv (preprint)*, 2025. doi: 10.1101/2025.09.23.25336497. doi:10.1101/2025.09.23.25336497.

[35] Fionn Büttner, Elaine Toomey, Sinead McClean, Mark Roe, and Eamonn Delahunt. Are questionable research practices facilitating new discoveries in sport and exercise medicine? the proportion of supported hypotheses is implausibly high. *British Journal of Sports Medicine*, 54(22):1365–1371, 2020. doi: 10.1136/bjsports-2019-101863.

[36] An-Wen Chan, Jennifer M. Tetzlaff, Peter C. Gøtzsche, Douglas G. Altman, Howard Mann, Jesse A. Berlin, et al. SPIRIT 2013 explanation and elaboration: Guidance for protocols of clinical trials. *BMJ*, 346:e7586, 2013. doi: 10.1136/bmj.e7586.

[37] Sean P. Riley, Brian T. Swanson, Stephen M. Shaffer, Steven F. Sawyer, and Joshua A. Cleland. The unknown prevalence of postrandomization bias in 15 physical therapy journals:

A methods review. *Journal of Orthopaedic & Sports Physical Therapy*, 51(11):542–550, 2021. doi: 10.2519/jospt.2021.10491.

[38] Nayara Santos Silva, Mark R. Elkins, Ítalo R. Lemes, Peter W. Stubbs, Márcia Rodrigues Franco, and Rafael Zambelli Pinto. Clinical trial registration has become more prevalent in physical therapy but it is still inadequate: A meta-research study. *Musculoskeletal Science and Practice*, 67:102854, 2023. doi: 10.1016/j.msksp.2023.102854.

[39] Brian J. McClenahan, Margaux Lojacono, Jodi L. Young, Ronald J. Schenk, and Daniel I. Rhon. Trials and tribulations of transparency related to inconsistencies between plan and conduct in peer-reviewed physiotherapy publications: A methodology review. *Journal of Evaluation in Clinical Practice*, 30(1):12–29, 2024. doi: 10.1111/jep.13810.

[40] Brian A. Nosek, George Alter, George C. Banks, Denny Borsboom, Sara D. Bowman, Steven J. Breckler, Stuart Buck, Christopher D. Chambers, et al. Promoting an open research culture. *Science*, 348(6242):1422–1425, 2015. doi: 10.1126/science.aab2374.

[41] Nórton Luís Oliveira, Cíntia Ehlers Botton, Angélica Trevisan De Nardi, and Daniel Umpierre. Methodological quality and reporting standards in systematic reviews with meta-analysis of physical activity studies: a report from the strengthening the evidence in exercise sciences initiative (sees initiative). *Systematic Reviews*, 10:304, 2021. doi: 10.1186/s13643-021-01845-9.

[42] Diane Rosen, Nils L. Reiter, Barbara Vogel, and Robert Prill. The use of systematic reviews for conducting new studies in physiotherapy research: a meta-research study comparing author guidelines of physiotherapy-related journals. *Systematic Reviews*, 13:28, 2024. doi: 10.1186/s13643-023-02427-7.

[43] Marco Bordino, Elisa Ravizzotti, and Stefano Vercelli. Retracted articles in rehabilitation: just the tip of the iceberg? a bibliometric analysis. *Archives of Physiotherapy*, 10:21, 2020. doi: 10.1186/s40945-020-00092-w.

[44] Claudio Bianchini, Carola Cosentino, Matteo Paci, and Marco Baccini. Open access physical therapy journals: Do predatory journals publish lower-quality randomized controlled trials? *Archives of Physical Medicine and Rehabilitation*, 101(6):969–977, 2020. doi: 10.1016/j.apmr.2019.12.012.

[45] Arianne P. Verhagen, Peter William Stubbs, Poonam Mehta, David Kennedy, Anthony M. Nasser, Camila Quel de Oliveira, Joshua W. Pate, and Ian W. Skinner. Comparison between 2000 and 2018 on the reporting of statistical significance and clinical relevance in physiotherapy clinical trials in six major physiotherapy journals: a meta-research design. *BMJ Open*, 12(1):e054875, 2022. doi: 10.1136/bmjopen-2021-054875.

[46] Nimish Mittal, Mohit Bhandari, and Dinesh Kumbhare. A tale of confusion from overlapping confidence intervals. *American Journal of Physical Medicine & Rehabilitation*, 98(1):81–83, 2019. doi: 10.1097/PHM.0000000000001016.

[47] Jorge Fuentes, Susan Armijo-Olivo, Bruno R. da Costa, Christine Ha, Humam Saltaji, Chiara Arienti, Stefano Negrini, and Greta G. Cummings. Does type of sponsorship of randomized controlled trials influence treatment effect size estimates in rehabilitation: A meta-epidemiological study. *American Journal of Physical Medicine & Rehabilitation*, 99 (10):909–916, 2020. doi: 10.1097/PHM.0000000000001444.

[48] Sathish Rajasekaran, Rodney Li Pi Shan, and Jonathan T. Finnoff. Honorary authorship: Frequency and associated factors in physical medicine and rehabilitation research articles. *Archives of Physical Medicine and Rehabilitation*, 95(3):418–428, 2014. doi: 10.1016/j.apmr.2013.09.024.

[49] James L. Nuzzo. Letters to the editor in exercise science and physical therapy journals: an examination of content and authorship inflation. *Scientometrics*, 126(8):6917–6936, 2021. doi: 10.1007/s11192-021-04068-w.

[50] Cameron McR. Gosling, Melainie Cameron, and Peter F. Gibbons. Referencing and quotation accuracy in four manual therapy journals. *Manual Therapy*, 9(1):36–40, 2004. doi: 10.1016/S1356-689X(03)00056-0.

[51] Matteo Paci, Nicola Landi, Giulia Briganti, and Bruno Lombardi. Factors associated with citation rate of randomised controlled trials in physiotherapy. *Archives of Physiotherapy*, 5: 9, 2015. doi: 10.1186/s40945-015-0009-6.

[52] Waqar M. Naqvi, Summaiya Zareen Shaikh, and Gaurav V. Mishra. Large language models in physical therapy: time to adapt and adept. *Frontiers in Public Health*, 12:1364660, 2024. doi: 10.3389/fpubh.2024.1364660.

[53] Douglas G. Altman. Comparability of randomised groups. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 34(1):125–136, 1985. doi: 10.2307/2987510.

[54] Stephen Senn. Testing for baseline balance in clinical trials. *Statistics in Medicine*, 13(17): 1715–1726, 1994. doi: 10.1002/sim.4780131703.

[55] Alessandro De Cassai, Burhan Dost, Esra Turunc, Engin İhsan Turan, Muzeyyen Beldagli, Mehmet Akif Yilmaz, Yunus Emre Karapinar, Annalisa Boscolo, and Paolo Navalesi. The enduring table 1 fallacy: A meta-research study of baseline testing in anesthesiology and pain trials. *Anesthesiology*, 144(1):156–162, 2025. doi: 10.1097/ALN.0000000000005776.

[56] Nathaniel Schenker and Jane F. Gentleman. On judging the significance of differences by examining the overlap between confidence intervals. *The American Statistician*, 55(3): 182–186, 2001. doi: 10.1198/000313001317097960.