Journal of Clinical Epidemiology 169 (2024) 111303

Journal of Clinical Epidemiology

ORIGINAL RESEARCH

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

Nicola Ferri[a,b,*], Elisa Ravizzotti[c], Alessandro Bracci[d], Giulia Carreras[e], Paolo Pillastrini[a,b], Mauro Di Bari[f,g]

[a]*Department of Biomedical and Neuromotor Sciences (DIBINEM), Alma Mater Studiorum University of Bologna, Bologna, Italy*
[b]*Division of Occupational Medicine, IRCCS University Hospital of Bologna S Orsola-Malpighi Polyclinic, Bologna, Italy*
[c]*Department of Neuroscience, Rehabilitation, Ophthalmology, Genetics and Maternal Child Health (DINOGMI), University of Genoa, Genoa, Italy*
[d]*Department for Life Quality Studies (QUVI), University of Bologna, Rimini, Italy*
[e]*Oncologic Network, Prevention and Research Institute (ISPRO), Florence, Italy*
[f]*Department of Experimental and Clinical Medicine, University of Florence, Florence, Italy*
[g]*Unit of Geriatrics, Department of Medicine and Geriatrics, Azienda Ospedaliero-Universitaria Careggi, Florence, Italy*

Accepted 19 February 2024; Published online 24 February 2024

Abstract

Objectives: To assess the confidence in the results of systematic reviews on the effectiveness of physiotherapy for musculoskeletal conditions in the past 10 years and to analyze trends and factors associated.

Methods: This is a metaepidemiological study on systematic reviews (SRs) with meta-analysis of randomized controlled trials (RCTs). MEDLINE, Cochrane Database of Systematic Reviews, CINAHL, and PEDro were searched for SRs of RCT on physiotherapy interventions for musculoskeletal disorders from December 2012 to December 2022. Two researchers independently screened the records based on the inclusion criteria; a random sample of 100 studies was selected, and each journal, author, and study variable was extracted. The methodological quality of SRs was independently assessed with the AMSTAR 2 tool. Any disagreement was solved by consensus.

Results: The confidence in SRs results was critically low in 90% of the studies, and it did not increase over time. Cochrane reviews are predominantly represented in the higher AMSTAR 2 confidence levels, with a statistically significant difference compared to non-Cochrane reviews. The last author's H-index is the only predictor of higher confidence among the variables analyzed (OR 1.04; 95% CI: 1.01, 1.06).

Conclusion: The confidence in SRs results is unacceptably low. Given the relevance of musculoskeletal disorders and the impact of evidence synthesis on the clinical decision-making process, there is an urgent need to improve the quality of secondary research by adopting more rigorous methods.

Keywords: Evidence-based practice; Musculoskeletal diseases; Physical and rehabilitation medicine; Systematic reviews as topic; Methods; Critical appraisal

Funding support: This research received no specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

* Corresponding author. Department of Biomedical and Neuromotor Sciences (DIBINEM), Alma Mater Studiorum - University of Bologna, via Pelagio Palagi 9, 40138, Bologna (BO), Italy.
E-mail address: nicola.ferri11@unibo.it (N. Ferri).

https://doi.org/10.1016/j.jclinepi.2024.111303

Plain language summary

The world population is aging, and muscle pain is the most frequent reason to see a physiotherapist. For the choice of treatment, systematic reviews are the best type of science to consider. In this study, we assessed the quality of research in this field. We found that most of them were of low quality, and this was constant over the years. The poor quality of these studies is an important problem for clinicians; indeed, they do not know what treatment is better after reading such papers. Science needs to improve and stress quality over quantity.

1. Introduction

Systematic reviews (SRs) with meta-analysis are essential to evidence-based medicine, as they are considered the best synthesis of intervention studies—in particular, of randomized controlled trials (RCT)—and a fundamental source of evidence for guideline developers and policymakers [1]. This type of study has been dramatically increasing in recent years, up to an estimate of 80 new SRs published daily [2].

Physiotherapy has clinical, social, and economic relevance within the scientific literature. It has been estimated that one-third of all the people in the world will need physiotherapy at some point in their lives [3]. Out of the variety of conditions that can benefit from physiotherapy, the most frequent indication to physiotherapy is represented by musculoskeletal disorders, which affect 1.71 billion people (95% CI: 1.68, 1.80) across the world, with low back pain being the leading condition [3].

Despite an increasing information overload, little is known about the overall quality of SRs and the impact of their results and conclusions on clinical practice. This can undermine the progress and credibility of research and be an obstacle to reducing the gap between researchers and clinicians. Recent metaresearch found that SRs were of low or critically low confidence in several diverse fields, such as treatments for Alzheimer's disease [4], adult major depression [5], mental and behavioral disorders [6], physical activity promotion [7], in vitro dental studies [8], childhood allergy [9], and surgical adverse events [10].

The "Publish or Perish" paradigm is one of the elements behind quantity over quality [11]. For this reason, there is still debate about the strengths and limitations of current bibliometrics in representing the quality and impact of scientific research in academics [12−15]. A recent survey found that more than 50% of decision-makers have difficulty choosing the best evidence on a given topic: the methodological quality, the reputation of the authors and the journal, and the type of primary studies included were among the features thought to be important by respondents [16].

The clinical need for high-quality evidence suffers from the general low SRs confidence; recently, researchers have developed methodologies to support clinical reasoning when there is uncertainty [17], new reporting guidelines for a better knowledge translation to the decision makers [18], and have suggested more careful quality assessments by editors and peer reviewers during submission [19].

On the other hand, with the aging of the world population, physiotherapy is essential to reduce the burden on health systems by contributing to prevention and health promotion, which are a consistent part of the clinical practice in physiotherapy. Thus, this study aimed to analyze the methodological quality of SRs of RCT on the effectiveness of physiotherapy for musculoskeletal conditions published in the last 10 years and to explore any characteristics associated with SR confidence.

2. Methods

2.1. Design

This is an observational study in the physiotherapy evidence synthesis field. Since there are no specific guidelines for metaepidemiological studies, and the MethodologIcal STudy reporting Checklist is still under development [20], we followed the Cochrane Handbook for guidance in the selection and extraction processes [21], and adhered to the Strengthening the Reporting of Observational Studies in Epidemiology guidelines [22]. This metaepidemiological study was prospectively registered on the Open Science Framework (https://osf.io/bc8zw/).

2.2. Eligibility criteria

2.2.1. Population

Our study focuses on musculoskeletal conditions defined and categorized by the International Classification of Diseases 11th Revision (ICD-11) [23].

2.2.2. Intervention

We included all types of treatment that physiotherapists in any part of the world can administer, considering the heterogeneity of norms and cultures; thus, dry needling, Tai Chi, and yoga were included, among other approaches. All studies on prevention in healthy subjects were excluded.

2.2.3. Control

We included all comparisons because our objective was to study the methodological quality, not to perform a synthesis of treatment effectiveness.

What is new?

Key findings

- Over the past 10 years, nine out of ten systematic reviews on musculoskeletal conditions had critically low confidence in results.
- In our sample, all systematic reviews with high and moderate confidence are Cochrane reviews; the h-index of the last author is the only predictor of high confidence.

What this adds to what was known?

- Our findings highlight important methodological issues that impact the optimal delivery of patient care and the liability of health-care professionals.

What is the implication and what should change now?

- Clinicians should carefully assess the quality of systematic reviews, regardless of any variable in the publication, journal, or study itself.
- Greater adherence to methodological guidelines and careful assessment during the peer-review process could improve confidence in systematic reviews.

2.2.4. Outcomes

For the same reason, we had no interest in considering only specific outcomes. The only restriction applied was to studies broadly addressing the clinical effectiveness of physiotherapy; thus, economic studies were excluded.

2.2.5. Study design

We included only SRs of RCT because they are considered the best summary of the available scientific evidence. This claim justifies their high impact on clinical practice in questions concerning treatment effectiveness. SRs incorporating also nonrandomized intervention studies were included only if a meta-analysis of RCT was reported separately. We excluded network meta-analyses, scoping reviews, rapid reviews, and other forms of evidence synthesis different from Cochrane's definition of SR [21]. We limited the sample to the last 10 years to avoid confounding factors related to a significant gap in the availability of methodological guidelines.

2.3. Searching

We searched MEDLINE (via PubMed), Cochrane Database of Systematic Reviews (via Cochrane Library), CINAHL (via EBSCOHost), and PEDro, from December 2012 to December 2022; the search strategy included both structured terms and free texts linked by logical operators (Appendix 1). Then, a semiautomated deduplication via EndNote (EndNote Web, Clarivate, Philadelphia, USA) was performed, followed by a manual check. One record was randomly selected if the same paper was published in multiple journals. The strategy for full-text retrieval included contacting the manuscript's authors when necessary.

2.4. Screening

Two researchers independently selected records by title and abstracts, based on the inclusion and exclusion criteria described above. The same process was repeated on the full texts retrieved. Any disagreement was solved by consensus. Then, we randomized all the full texts using an Excel function (Office 2021, Microsoft, Redmond, WA, USA), and the first 100 SRs were selected for the purposes of our study.

2.5. Extraction

A researcher extracted data of interest (Table 1) from the SRs into a predefined spreadsheet; a second reviewer then checked all the extractions. Conflicts were solved by consensus.

Data extracted were coded as follows.

- Type of SRs identifies Cochrane vs non-Cochrane review.
- The Journal's impact factor (JIF) was assessed at publication date via the Clarivate Journal Citation Report (JCR)
- The Journal's Quartile was assessed at publication date via Scimago and Clarivate. When a journal was classified as belonging to different discipline areas, the closer to physiotherapy was considered.
- The first and last authors' H-indexes were assessed via Scopus, including documents until the publication year of the SR considered.
- The publication policy was determined by consulting the Directory of Open Access Journals, JCR, Scopus, the journals' website, and the free full-text availability.
- The total studies included refers to the papers included in the SRs, not in meta-analyses.
- Results and Conclusions were extracted referring to the primary outcome; if more than 1 primary outcome was present or no primary outcome was mentioned, we extracted the first outcome that was reported. We considered the Results favorable if statistical significance was obtained and the conclusions favorable if a recommendation toward the intervention was given.
- Country refers to the corresponding author's primary affiliation; we categorized these data into the continent for statistical analysis.

Table 1. Data extracted

General variables	Authors variables	Journal variables	Study variables
First author	Country	JIF	Type of SRs
Year of publication	Number of authors	Journal Q (Scimago)	Publication policy
Study title	First author H-index	Journal Q (JCR)	Total studies included
Journal name	Last author H-index		Study design included
			Results
			Conclusions

JCR, Journal Citation Report; JIF, Journal Impact Factor; Q, Quartile; SRs, Systematic Review with Meta-Analysis.

2.6. A Measurement Tool to Assess Systematic Reviews—version 2 (AMSTAR 2) assessment

AMSTAR 2 is a critical appraisal tool for systematic reviews, consisting of 16 items on methodological domains with *yes/no* and *partial yes* ratings, with items 2, 4, 7, 9, 11, 13, and 15 referring to critical domains [24]. The psychometric characteristics of AMSTAR 2 have been studied and validated, comparing it with pre-existing tools [25].

Three researchers performed deep training following the AMSTAR 2 publication, guidance, and online educational support. A piloting test was completed before the assessments, and there was a discussion with AMSTAR 2's

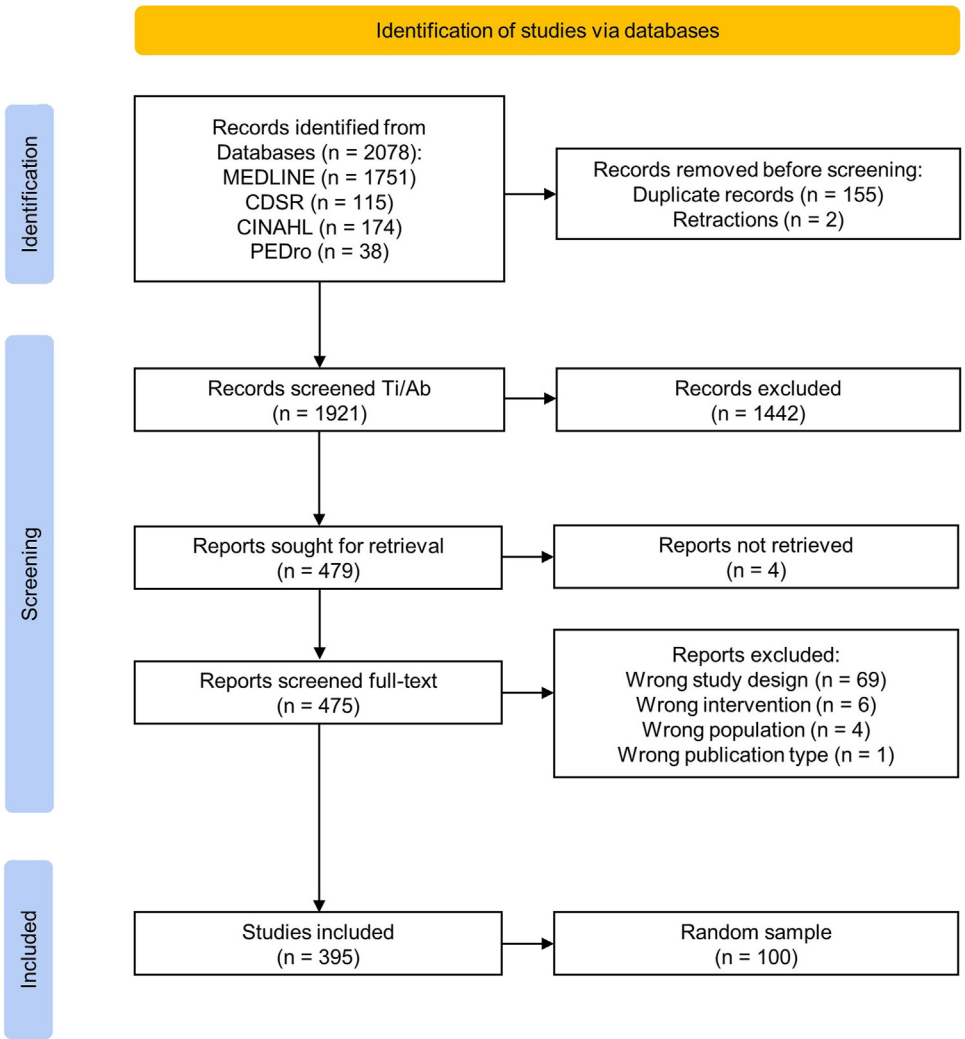

Figure 1. Study flowchart. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

authors on some judgment calls to enhance the most accurate interpretation and consistency. Then, 2 independent researchers assessed each SR included, and any conflict was solved by consensus. A summary score was not calculated; whereas, as Shea et al. [24] recommended, four confidence levels (ie, critically low, low, moderate, and high) were identified.

2.7. Data analysis

Data were analyzed with descriptive statistics: we used mean and SD for continuous data (or median and interquartile range for non-normal distribution), counts and percentages for categorical data. All the variables were grouped according to the AMSTAR 2 score, and statistical tests were performed to analyze any significant differences. For comparing the variables' distribution among the AMSTAR 2 categories the Kruskal−Wallis, chi-squared test or 1-way ANOVA tests were used depending on the variable's nature. To explore for confounding in the association between JIF and confidence ratings, we performed a sensitivity analysis excluding Cochrane SRs.

To preserve the ordinal distribution of the four AMSTAR 2 levels, an ordinal logistic regression analysis model was built to identify predictors of the overall SRs confidence among the variables extracted. The dependent variable was categorized by merging two classes in the three categories critically low; low; moderate-high. To also test the AMSTAR 2 publication as a predictor of SR confidence, we dichotomized the year of publication of the SRs included (<2018 vs ≥ 2018), considering a reasonable uptake period for the AMSTAR 2 tool. Statistical analyses were performed with Stata 17 software (StataCorp. 2021, College Station, TX, USA).

3. Results

The search strategy retrieved 2078 records, of which 155 were duplicates and 2 were retracted papers. Thus, 1921 records were screened by title and abstract, leading to 475 full texts assessed for eligibility. Finally, 395 SRs were included (Fig 1), of which a random sample of 100 was obtained. We reported details of the studies included (Appendix 2) and the reasons for full-text exclusions (Appendix 3).

Our sample includes different types of interventions currently used in musculoskeletal physiotherapy around the world, with a major prevalence of manual therapy, exercise, physical agents, or a combination thereof (Fig 2).

Overall, 90% of the studies selected have critically low confidence in the results. Moreover, the methodological quality of the studies did not increase over time (Table 2).

After the exclusion of eight journals that had no JIF at the publication date, the JIF ranged from 0.445 to 30.313, and the differences between the AMSTAR 2 levels were statistically significant (Table 3). However, there is no association between the two variables if the Cochrane SRs are excluded from the analysis (Appendix 4).

As shown in Table 4, every continent was represented in our sample. The H-index of the first and last authors ranged from 0 to 39 and 1 to 100, respectively; these variables had a significantly different distribution according to the

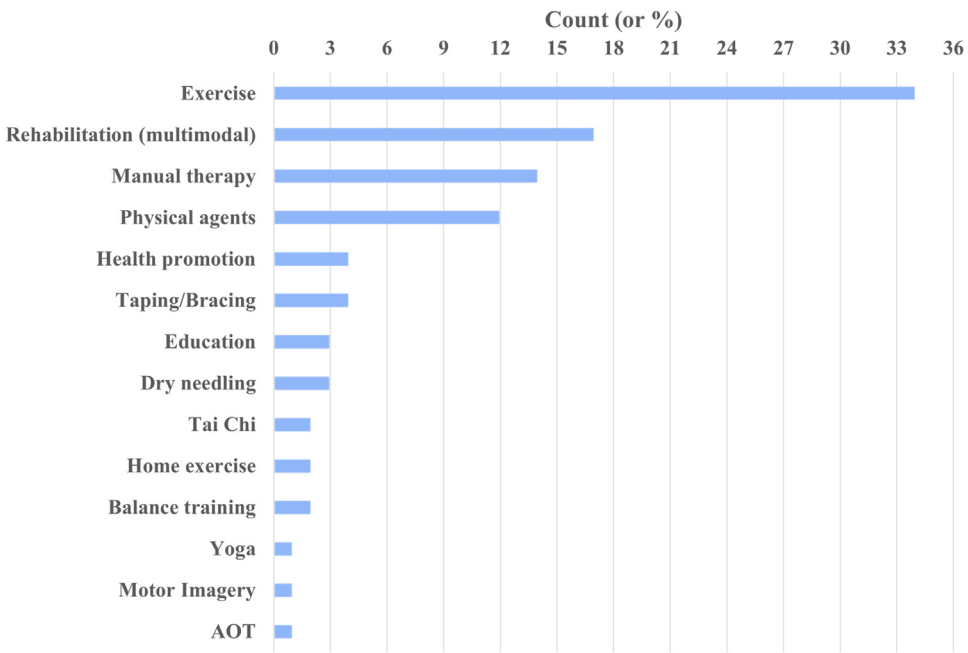

Figure 2. Type of intervention included. Abbreviations: AOT, Action Observation Therapy. Multimodal means the combination of manual therapy, exercise, or physical agents. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 2. SRs confidence over time

		AMSTAR 2				
		Critically low	Low	Moderate	High	
Independent variable	Total	90	4	2	4	*P* value
Year of publication, *n* (%)						0.200[a]
2013	4	3 (75)	0 (0)	1 (25)	0 (0)	
2014	4	3 (75)	0 (0)	0 (0)	1 (25)	
2015	5	5 (100)	0 (0)	0 (0)	0 (0)	
2016	10	9 (90)	0 (0)	1 (10)	0 (0)	
2017	12	11 (91.7)	0 (0)	0 (0)	1 (8.3)	
2018	12	10 (83.3)	1 (8.3)	0 (0)	1 (8.3)	
2019	13	11 (84.6)	1 (7.7)	0 (0)	1 (7.7)	
2020	12	10 (83.3)	2 (16.7)	0 (0)	0 (0)	
2021	13	13 (100)	0 (0)	0 (0)	0 (0)	
2022	15	15 (100)	0 (0)	0 (0)	0 (0)	

AMSTAR 2, a measurement tool to assess systematic reviews—version 2; SR, systematic review.
[a] Chi-squared test.

AMSTAR 2 ratings. The only moderate or high confidence SRs were Cochrane reviews.

Analyzing the single items (Fig 3), 93% of the studies did not explain the reason for the eligibility criteria of study designs (item 3), 78% did not report the list of the studies excluded (item 7), and 90% did not check the funding sources of the primary studies included (item 10); also, less than 10% of the SRs had a comprehensive search strategy (item 4), which is a critical domain of the AMSTAR 2 tool. The strengths of the SRs are accurate PICO reporting (89%), declaration of conflicts of interest (93%), and good methods for selection process (88%).

3.1. Ordered logistic regression

An ordered logistic regression model for analyzing AMSTAR 2 scores was built (Table 5). Covariates that satisfied the proportional odds assumption were identified and the best fitting model in terms of Akaike Information Criterium was selected. Results are displayed as proportional odds ratios.

The JIF was not a predictor of SR confidence. Conversely, the odds of a high AMSTAR 2 rating vs the combined critically low, low, and moderate AMSTAR 2 ratings were 1.04 times greater for each unit increase in the H-index of the last author, assuming that all other variables in the model were held constant. Because of the proportional odds assumption, the same increase of 1.04 times would represent the risk of being in the combined AMSTAR 2 high and moderate categories vs the critically low and low ones. Increasing the number of authors by one resulted in 53% higher odds for high AMSTAR 2 than the other categories, with a borderline significant association. Cochrane SRs were associated with confidence

Table 3. Journal variables

		AMSTAR 2				
		Critically low	Low	Moderate	High	
Independent variables	Total	90	4	2	4	*P* value
JIF, median (IQR)	2.99 (2.36–4.77)	2.90 (2.18–4.06)	4.14 (3.09–17.55)	6.10 (5.94–6.26)	7.25 (6.39–7.82)	0.004[b]
Journal Q (Scimago), *n* (%)						0.828[a]
1	68	58 (85.3)	4 (5.9)	2 (2.9)	4 (5.9)	
2	21	21 (100)	0 (0)	0 (0)	0 (0)	
3	9	9 (100)	0 (0)	0 (0)	0 (0)	
4	1	1 (100)	0 (0)	0 (0)	0 (0)	
Journal Q (JCR), *n* (%)						0.668[a]
1	50	41 (82)	3 (6)	2 (4)	4 (8)	
2	26	25 (96.1)	1 (3.9)	0 (0)	0 (0)	
3	11	11 (100)	0 (0)	0 (0)	0 (0)	
4	5	5 (100)	0 (0)	0 (0)	0 (0)	

JCR, journal citation report; JIF, journal impact factor; Q, quartile.
[a] Chi-squared test.
[b] Kruskal–Wallis test.

Table 4. Authors and study variables

		AMSTAR 2				
		Critically low	Low	Moderate	High	
Independent variables	**Total**	**90**	**4**	**2**	**4**	***P* value**
Country, *n* (%)						0.935[a]
Africa	1	1 (100)	0 (0)	0 (0)	0 (0)	
Asia	27	26 (96.3)	1 (3.7)	0 (0)	0 (0)	
Central America	2	2 (100)	0 (0)	0 (0)	0 (0)	
Europe	34	29 (85.3)	2 (5.9)	2 (5.9)	1 (2.9)	
North America	6	5 (83.3)	0 (0)	0 (0)	1 (16.7)	
Oceania	14	13 (92.9)	0 (0)	0 (0)	1 (7.1)	
South America	14	12 (85.7)	1 (7.1)	0 (0)	1 (7.1)	
Number of authors, mean (SD)	5.56 (1.80)	5.42 (1.82)	6.50 (1.00)	7.5 (0.71)	6.75 (0.96)	0.326[c]
First author H-index, median (IQR)	4.50 (2.00–8.00)	4.00 (2.00–8.00)	16.50 (9.50–24.00)	18.50 (2.00–35.00)	6.50 (5.00–8.00)	0.045[b]
Last author H-index, median (IQR)	16.50 (8.00–30.50)	14.50 (7.00–28.00)	49.00 (26.00–84.50)	33.50 (29.00–38.00)	31.50 (22.50–46.00)	0.014[b]
Type of SRs, *n* (%)						<0.001[a]
non-Cochrane	90	86 (95.6)	4 (4.4)	0 (0)	0 (0)	
Cochrane	10	4 (40)	0 (0)	2 (20)	4 (40)	
Publication policy, *n* (%)						0.867[a]
non–open Access	62	55 (88.7)	3 (4.8)	1 (1.6)	3 (4.8)	
Open Access	38	35 (92.1)	1 (2.6)	1 (2.6)	1 (2.6)	
Total studies included, median (IQR)	11.50 (7.00–19.50)	12.00 (7.00–20.00)	15.00 (7.5–33.00)	8.00 (3.00–13.00)	8.00 (6.50–19.00)	0.677[b]
Study design included, *n* (%)						0.083[a]
RCTs and non-RCTs	15	12 (80)	2 (13.3)	1 (6.7)	0 (0)	
Only RCT	85	78 (91.8)	2 (2.4)	1 (1.2)	4 (4.7)	
Results, *n* (%)						0.644[a]
Not favorable	40	36 (90)	2 (5)	0 (0)	2 (5)	
Favorable	60	54 (90)	2 (3.3)	2 (3.3)	2 (3.3)	
Conclusions, *n* (%)						0.184[a]
Not favorable	33	28 (84.8)	2 (6.1)	0 (0)	3 (9.1)	
Favorable	67	62 (92.5)	2 (3)	2 (3)	1 (1.5)	

RCT, randomized controlled trial; SR, Systematic Review with Meta-Analysis.
[a] Chi-squared test.
[b] Kruskal–Wallis test.
[c] 1-way ANOVA test.

because, compared to non-Cochrane ones, they were predominantly represented in the higher AMSTAR 2 ratings. However, the variable violated the assumption of proportionality and did not remain associated with the outcome in multivariate comparisons. The variable of the year of AMSTAR 2 introduction does not satisfy the proportional assumption, and the model that incorporates such variable performs worse than the final model. In addition, it was not a predictor of SR confidence: the odds of a high AMSTAR 2 rating vs the combined critically low, low, and moderate AMSTAR 2 ratings was 18% lower (OR = 0.82, 95% CI: 0.19, 3.50) for studies published in 2018 or after in comparison to those published before 2018, assuming that all other variables in the model were held constant.

4. Discussion

In this representative sample of SRs on musculoskeletal physiotherapy published in the last 10 years, the SR confidence was critically low in 90% of the studies retrieved and

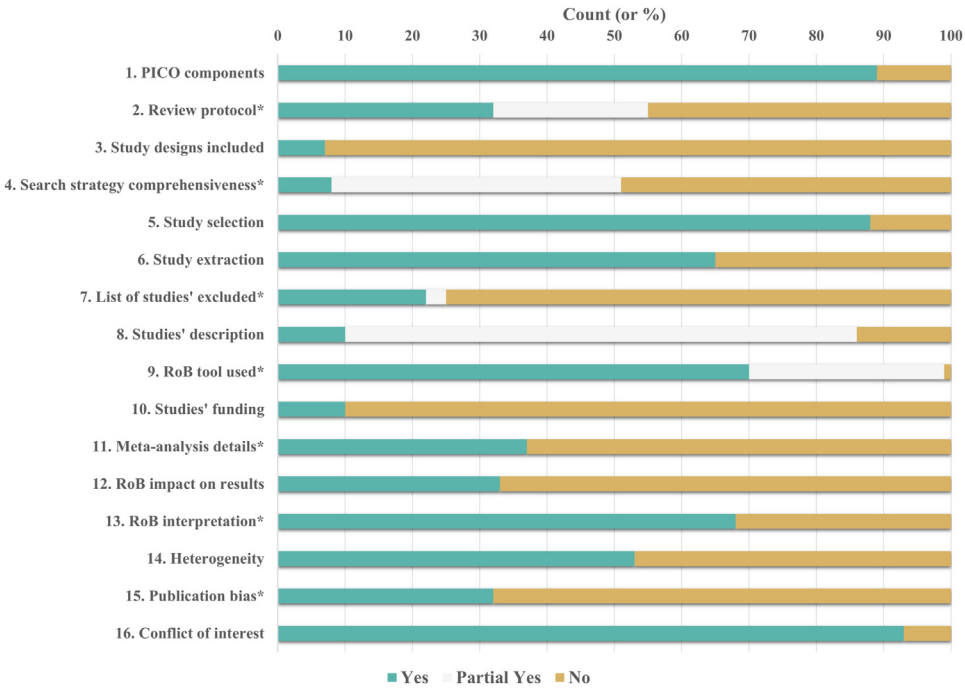

Figure 3. AMSTAR 2 items assessment. * Critical domains. AMSTAR 2 = A Measurement Tool to Assess Systematic Reviews—version 2. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

did not show any time-dependent increase. Of the candidate predictors, only the last author's H-index predicted the SRs confidence.

The practice of health-care professions implies an obligation of means toward the patient that translates into acting with prudence and diligence. To achieve this, health-care providers should apply the best scientific evidence available for clinical and legal purposes. Yet, what if the evidence is of low quality?

This metaresearch strengthens the findings of Riley et al. [26], who reported low confidence on the results of 24 SRs on a similar topic. Assessing methodological quality could be challenging since the AMSTAR 2 tool draws heavily on reporting; our difficulty is supported by a recent study that found a significant association between the quality of reporting and the risk of bias of SRs [27]. In our study, these challenges were addressed through preliminary extensive training and piloting, and the consensus-based decision-making process. Retrospectively, our choices were consistent with recent recommendations [28]. We also found a tendency of AMSTAR 2 to a floor effect, as described in previous research [29], which, together with our rigorous approach, could explain low confidence attributed to some less recent Cochrane reviews. Thus, one might wonder whether the SRs were of low confidence or the tool was too rigid and demanding. Perhaps considering item 7 as a critical domain could be severe compared to the other critical items: authors reasonably have a list of excluded studies with justifications but may have chosen not to report it. Interestingly, this item would be fulfilled by strict adherence to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement (item 16b). Finally, a high standard is definitely desirable, and besides, we did find some high-confidence reviews, all from Cochrane Collaboration. This finding indicates that such methodological quality levels can be achieved; thus, methodological improvement is still a priority.

Table 5. Multivariable ordinal logistic regression predicting the confidence of the results of 100 systematic reviews with meta-analysis of randomized controlled trials on musculoskeletal physiotherapy

Variable	OR (95% CI)
JIF	1.00 (0.99, 1.01)
Last author H-index	1.04 (1.01, 1.06)
Number of authors	1.53 (0.99, 2.39)
Intercept 1	5.8
Intercept 2	6.44
Intercept 3	6.88

CI, confidence interval; JIF, journal impact factor; OR, odds ratio.

The last author H-index was the only variable with a predictive value for SR confidence. This finding is consistent with the last author being usually a senior researcher, responsible for the choice and training of collaborators and the overall scientific architecture of the study. In a way, it also represents a small comfort in the goodness of the index, knowing all its recent limitations and criticalities [12–15].

The main strengths of this study are the comprehensive and rigorous selection process, the use of a recognized, standardized tool for quality assessment, and the statistical approach that accurately reflects the ordinal nature of the

outcome variable. At the same time, some limitations must be acknowledged. First, in the expectation of a possible time-dependent trend in methodological quality, we included a broad and representative number of studies published in the last 10 years; however, the AMSTAR 2 tool was published only in 2017, more recently than the oldest studies we selected. Nevertheless, the tool was designed to assess methodological quality, not as a guidance to the SR conduction; moreover, the previous version AMSTAR and the Cochrane Handbook were both already available before 2017 [30,31]. Years of publication trended to be negatively correlated with SR confidence, although this variable could not be included in the multivariable ordinal model due to a violation of the proportionality assumption. Interestingly, a recent metaresearch study found that 42 out of 43 SRs adhering to AMSTAR 2 were still of low or critically low confidence [19]. Second, to optimize resources, we did not analyze the entire sample of 395 SRs but only a share of them. Because the sample was drawn randomly and included more than a quarter of the studies retrieved, we are confident of its representativeness and, consequently, of the validity of our results, which are consistent with previous studies. Third, considering the study design, our search strategy had to be specific enough, using MeSH terms and free text consistent with the ICD-11 musculoskeletal disorders definition. It is possible that we missed some studies, but post hoc, our sample is representative of the population and well distributed with respect to the variables analyzed. Fourth, although we found a predictive value of last author H-index, it should be pointed out that it is a highly dynamic metric subject to bias. Fifth, the regression analysis suffers from data imbalance classifiable as a rare event problem that leads to biased estimates. Imbalanced problems have been widely studied for standard classification problems such as applying penalized likelihood, but in the case of ordinal data, alternative methods such as active learning or neural network algorithms are still under study [32,33]. Last, the open-access variable was attributed to each specific article; we realize that this dichotomization does not fully capture the heterogeneity of publication policies and may result in less sensitive data.

5. Conclusion

Evidence-based medicine involves considering patient expectations and seeking the best scientific evidence to support clinical reasoning and decision-making. This research alerts us to the very low confidence of SRs on treatment effectiveness for such a largely prevalent condition as musculoskeletal disorders, representing the first indication for physiotherapy in the world. As a result, our findings may directly impact the delivery of optimal care to patients and the health-care providers liability. Better adherence to current scientific methods, including reporting guidelines, is imperative to achieve higher quality in scientific research. Enhancing prepublication peer review processes could be a possible effective strategy.

CRediT authorship contribution statement

Nicola Ferri: Writing − review & editing, Writing − original draft, Methodology, Investigation, Data curation, Conceptualization. **Elisa Ravizzotti:** Writing − review & editing, Investigation, Data curation, Conceptualization. **Alessandro Bracci:** Writing − review & editing, Investigation, Data curation, Conceptualization. **Giulia Carreras:** Writing − review & editing, Methodology, Formal analysis, Data curation, Conceptualization. **Paolo Pillastrini:** Writing − review & editing, Supervision, Conceptualization. **Mauro Di Bari:** Writing − review & editing, Writing − original draft, Supervision, Methodology, Formal analysis, Conceptualization.

Data availability

All data supporting the findings of this study are available within the paper and the Supplementary materials.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We are grateful to Dr Beverly Shea for her feedback on some AMSTAR 2 tool's details.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jclinepi.2024.111303.

References

[1] Clarke J. What is a systematic review? Evid Based Nurs 2011;14(3):64.
[2] Hoffmann F, Allers K, Rombey T, Helbach J, Hoffmann A, Mathes T, et al. Nearly 80 systematic reviews were published each day: observational study on trends in epidemiology and reporting over the years 2000-2019. J Clin Epidemiol 2021;138:1−11.
[3] Cieza A, Causey K, Kamenov K, Hanson SW, Chatterji S, Vos T. Global estimates of the need for rehabilitation based on the global burden of disease study 2019: a systematic analysis for the global burden of disease study 2019. Lancet Lond Engl 2021;396(10267):2006−17.
[4] Zhong CCW, Zhao J, Wong CHL, Wu IXY, Mao C, Yeung JWF, et al. Methodological quality of systematic reviews on treatments for

Alzheimer's disease: a cross-sectional study. Alzheimer's Res Ther 2022;14(1):159.
[5] Matthias K, Rissling O, Pieper D, Morche J, Nocon M, Jacobs A, et al. The methodological quality of systematic reviews on the treatment of adult major depression needs improvement according to AMSTAR 2: a cross-sectional study. Heliyon 2020;6(9):e04776.
[6] De Santis KK, Lorenz RC, Lakeberg M, Matthias K. The application of AMSTAR2 in 32 overviews of systematic reviews of interventions for mental and behavioural disorders: a cross-sectional study. Res Synth Methods 2022;3(4):424−33.
[7] De Santis KK, Matthias K. Different approaches to appraising systematic reviews of digital interventions for physical activity promotion using AMSTAR 2 tool: cross-sectional study. Int J Environ Res Public Health 2023;20(6):4689.
[8] Hammel C, Pandis N, Pieper D, Faggion CM Jr. Methodological assessment of systematic reviews of in-vitro dental studies. BMC Med Res Methodol 2022;22:110.
[9] Matterne U, Theurich MA, Pröbstl S, Pieper D, Wang J, Xu A, et al. Quality of systematic reviews on timing of complementary feeding for early childhood allergy prevention. BMC Med Res Methodol 2023;23:80.
[10] Zhou X, Li L, Lin L, Ju K, Kwong JSW, Xu C. Methodological quality for systematic reviews of adverse events with surgical interventions: a cross-sectional survey. BMC Med Res Methodol 2021;21:223.
[11] Siegel MG, Brand JC, Rossi MJ, Lubowitz JH. "Publish or perish" promotes medical literature quantity over quality. Arthroscopy 2018;34(11):2941−2.
[12] Curry S. Let's move beyond the rhetoric: it's time to change how we judge research. Nature 2018;554:147.
[13] Hunter PR. Bibliometrics, research quality, and neglected tropical diseases. Lancet 2009;373(9664):630−1.
[14] Singh HP. Alternative research bibliometrics: it's about quality and not quantity. Shoulder Elbow 2022;14(2):121−2.
[15] Weightman AL, Butler CC. Using bibliometrics to define the quality of primary care research. BMJ 2011;342:d1083.
[16] Lunny C, Whitelaw S, Reid EK, Chi Y, Zhang JH, Ferri N, et al. Can artificial intelligence help decision makers navigate the growing body of systematic review evidence? A cross-sectional survey, 10 January 2024, PREPRINT (Version 2) available at Research Square. Available at: https://www.researchsquare.com/article/rs-2416773/v2. Accessed March 9, 2024.
[17] Gattrell WT, Logullo P, van Zuuren EJ, Price A, Hughes EL, Blazey P, et al. ACCORD (ACcurate COnsensus Reporting Document): a reporting guideline for consensus methods in biomedicine developed via a modified Delphi. PLoS Med 2024;21(1):e1004326.
[18] Gates M, Gates A, Pieper D, Fernandes RM, Tricco AC, Moher D, et al. Reporting guideline for overviews of reviews of healthcare interventions: development of the PRIOR statement. BMJ 2022;378:e070849.
[19] Bojcic R, Todoric M, Puljak L. Most systematic reviews reporting adherence to AMSTAR 2 had critically low methodological quality: a cross-sectional meta-research study. J Clin Epidemiol 2024;165:111210.
[20] Lawson DO, Puljak L, Pieper D, Schandelmaier S, Collins GS, Brignardello-Petersen R, et al. Reporting of methodological studies in health research: a protocol for the development of the Methodological STudy reportIng Checklist (MISTIC). BMJ Open 2020;10(12):e040478.
[21] Cochrane handbook for systematic reviews of interventions. Available at: https://training.cochrane.org/handbook/current. Accessed August 10, 2023.
[22] von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. PLoS Med 2007;4(10):e296.
[23] ICD-11. Available at: https://icd.who.int/en. Accessed July 4, 2023.
[24] Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. BMJ 2017;358:j4008.
[25] Lorenz RC, Matthias K, Pieper D, Wegewitz U, Morche J, Nocon M, et al. A psychometric study found AMSTAR 2 to be a valid and moderately reliable appraisal tool. J Clin Epidemiol 2019;114:133−40.
[26] Riley S, Swanson BT, Sawyer SF, Brismée JM, Staysniak G. Should low-quality evidence dominate high-level evidence? A systematic review and meta-analysis of systematic reviews of musculoskeletal physical therapy interventions. J Man Manip Ther 2021;29(4):203−15.
[27] Innocenti T, Feller D, Giagio S, Salvioli S, Minnucci S, Brindisino F, et al. Adherence to the PRISMA statement and its association with risk of bias in systematic reviews published in rehabilitation journals: a meta-research study. Braz J Phys Ther 2022;26(5):100450.
[28] De Santis KK, Pieper D, Lorenz RC, Wegewitz U, Siemens W, Matthias K. User experience of applying AMSTAR 2 to appraise systematic reviews of healthcare interventions: a commentary. BMC Med Res Methodol 2023;23:63.
[29] Lorenz RC, Matthias K, Pieper D, Wegewitz U, Morche J, Nocon M, et al. AMSTAR 2 overall confidence rating: lacking discriminating capacity or requirement of high methodological quality? J Clin Epidemiol 2020;119:142−4.
[30] Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Med Res Methodol 2007;7:10.
[31] Higgins JPT, Green S. Cochrane Handbook for systematic reviews of interventions. Version 5.1.0.. The Cochrane Collaboration; 2011. Available at: www.handbook.cochrane.org. Accessed July 4, 2023.
[32] Ge J, Chen H, Zhang D, Hou X, Yuan L. Active learning for imbalanced ordinal regression. IEEE Access 2020;8:180608−17.
[33] Lázaro M, Figueiras-Vidal AR. Neural network for ordinal classification of imbalanced data by minimizing a Bayesian cost. Pattern Recogn 2023;137:109303.