Original Research

Conclusiveness of Cochrane systematic reviews is low but accumulating across time in physical therapy: A meta-research study

Ana Paula Coelho Figueira Freire [a,*] [ID], Kailei J. Eustis [a] [ID], Ítalo Ribeiro Lemes [b,c] [ID],
Bruno Tirotti Saragiotto [d], Gordon Smilanich [a], Mark Elkins [e], Rafael Zambelli Pinto [d,f]

[a] Department of Health Sciences, Central Washington University, Ellensburg, WA, USA
[b] Faculty of Science and Technology, Sao Paulo State University (UNESP), São Paulo, Brazil
[c] Faculdade Israelita de Ciências da Saúde Albert Einstein (FICSAE), São Paulo, Brazil
[d] Discipline of Physiotherapy, Graduate School of Health, Faculty of Health, University of Technology Sydney, New South Wales, Australia
[e] Faculty of Medicine and Health, University of Sydney, Sydney, Australia
[f] Department of Physical Therapy, Universidade Federal de Minas Gerais (UFMG), Minas Gerais, Brazil

ABSTRACT

Background: Conclusiveness of a systematic review is the ability to reach a definitive conclusion about the effectiveness of one treatment compared to another. The large proportion of systematic reviews that do not reach a definitive conclusion might discourage clinicians from engaging with and interpreting systematic reviews.
Objectives: To determine the percentage of conclusive Cochrane reviews in physical therapy and to investigate whether this percentage has increased over time.
Methods: In this meta-research study, we performed a systematic search of the Physiotherapy Evidence Database (PEDro) for Cochrane reviews. We extracted a random sample of 200 published systematic reviews, with 50 reviews from each of the periods: 2001 to 2005, 2006 to 2010, 2011 to 2015, and 2016 to 2020. Two independent assessors extracted information. The Grading of Recommendations Assessment, Development and Evaluation (GRADE) data for the primary outcomes was used to assess conclusiveness. Reviews were considered conclusive when at least one primary outcome provided high certainty of evidence.
Results: Outcomes with very low certainty of evidence represented 21 % of outcomes and increased from 14 % in 2001–2005 to 34 % in 2016–2020. Outcomes with low certainty of evidence comprised 55 % of outcomes and remained consistent over time. Moderate- and high-certainty outcomes remained consistent, composing 22 % and 2 % of outcomes, respectively. The proportion of outcomes with high certainty of evidence never exceeded 4 % per period. The percentage of conclusive reviews remained unchanged and consisted of 3 % of reviews in the sample; however, because the total number of reviews is increasing, there has been an accumulation in the number of conclusive reviews across the 20-year period.
Conclusions: The percentage of Cochrane reviews deemed conclusive remains small, although conclusive reviews are accumulating over time.

Introduction

High-quality systematic reviews are relevant to decision-making in healthcare and health policies.[1] The Cochrane Collaboration has played a unique role in the development of methodology for systematic reviews throughout its history and is recognized as representing an international gold standard for high-quality systematic reviews.[2,3] Cochrane systematic reviews tend to be of higher quality, are less vulnerable to bias, and

acknowledge more limitations, promoting greater transparency in identifying and reporting potential biases within the reviewed evidence. In addition, these reviews are generally more conservative in how results are endorsed than non-Cochrane reviews.[2-4]

Although Cochrane reviews provide a more accurate estimate of the effects of an intervention, there is still a debate among scientists and clinicians regarding their conclusiveness. Conclusiveness of a systematic review has been defined as the ability to reach a practical, definitive

conclusion about the effectiveness of one treatment compared to another treatment or placebo.[5] Previous studies showed that a considerable proportion of Cochrane reviews are inconclusive and highlighted the need for further and better studies to reduce uncertainty.[6,7] Inconclusiveness rates in Cochrane reviews may vary according to the field of study, from 20 % in pediatric gastroenterology[8] to 55 % in palliative and supportive care for cancer.[7] In physical therapy, 94.3 % of Cochrane reviews have been considered inconclusive.[6] These results may be partly due to the challenges of conducting high-quality controlled trials in this field.[9] Despite these difficulties, the number of physical therapy trials is accumulating exponentially and the quality of physical therapy trials is improving over time[10] so the landscape in which Cochrane reviews are conducted continues to evolve.

Cochrane reviews focus on synthesizing high-quality evidence, and although they are not intended to provide specific recommendations for clinical practice, clinicians often perceive a lack of clinical conclusiveness in systematic reviews, or even the absence of straightforward recommendations (i.e., reviews that do not offer a conclusion due to insufficient evidence). This can be a barrier to clinicians attempting to apply evidence generated by Cochrane reviews.[11-13] The large proportion of systematic reviews that do not reach a practical, definitive conclusion might discourage some clinicians from developing skills for interpreting systematic reviews. Therefore, it becomes relevant to investigate the prevalence of conclusiveness of systematic reviews and whether it is improving over time.

To our knowledge, only one study investigated the conclusiveness of Cochrane systematic reviews in physical therapy.[6] Despite the high prevalence of inconclusive Cochrane reviews, the search period was limited to reviews published between 2008 and 2017. In addition, the authors considered a review to be conclusive about the primary outcome if one intervention was superior to the alternative (control), or if the interventions were equivalent. We would argue that this definition is not accurate and does not account for contemporary methods to rate the overall certainty of evidence, such as the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach.[14] Therefore, the aims of this meta-research study were to investigate: (I) the proportion of conclusive Cochrane reviews in physical therapy and whether the level of conclusiveness has changed over time, and (II) the proportion of reviews reporting "need for further studies" and whether this proportion has changed over time.

Methods

Information source and search strategy

We performed a systematic search in the Physiotherapy Evidence Database (PEDro) for Cochrane reviews. To be indexed on PEDro, a systematic review must (i) contain a methods section that describes the search strategy and inclusion criteria; (ii) include at least one trial, review, or guideline (or explicitly search for but not find a trial, review, or guideline) that satisfies the criteria for inclusion in PEDro; and (iii) be a full paper (not an abstract) in a peer-reviewed journal. Two PEDro staff independently assess each new monthly issue of the Cochrane Database of Systematic Reviews for systematic reviews relevant to the field of physical therapy. This procedure allows all relevant Cochrane systematic reviews to be indexed on PEDro as soon as they are published.

We chose to use PEDro because this database exclusively indexes systematic reviews and trials relevant to physical therapy, ensuring the comprehensive inclusion of Cochrane reviews directly applicable to the physical therapy field and our research question.

We extracted a random sample of 200 published Cochrane systematic reviews, with 50 reviews from each of four periods: 2001 to 2005, 2006 to 2010, 2011 to 2015, and 2016 to 2020. The random sampling was performed using Microsoft Excel software (Microsoft Office 2007, Microsoft Corporation, Redmond, Washington), which allowed us to include representative samples at regular time points for analyzing

changes in the proportion of conclusive reviews over time.

The selection of the years was made following pilot searches conducted at the beginning of the project, which revealed a very limited number of physical therapy-relevant reviews published before 2001. Therefore, we divided the 20-year period into four equal intervals (2001–2005, 2006–2010, 2011–2015, and 2016–2020) to ensure consistency and systematic sampling.

The sample size of 200 reviews was based on similar meta-research studies examining studies indexed on PEDro.[15-17] These random samples have been deemed representative and can be extrapolated to draw implications about the larger cohorts of studies from which they were randomly sampled.

Selection of reviews

Any Cochrane review indexed on the PEDro database was considered eligible for this study. Cochrane reviews that have been withdrawn and Cochrane review protocols were not considered for this study. Cochrane reviews that included zero studies were excluded.

Data extraction

Two independent assessors extracted the following data from the included reviews: year of publication, country of first author, Cochrane Review Group, participants, intervention, comparators, primary outcomes, number of included randomized controlled trials, total cumulative number of patients enrolled, and the reporting of need for further studies. Reviews were categorized into subdisciplines based on those listed in PEDro using a combination of review titles and Cochrane Review Group affiliation. Participants were classified by the presence of acute or chronic conditions, or age if deemed healthy by the review. Participants grouped into the "other" category spanned multiple age groups and/or fell into a category that did not distinguish between healthy and unhealthy individuals (e.g., "pregnant women"). Intervention categories were created if they appeared more than once in the sample.

Assessors identified all primary outcomes in each review. For the first three primary outcomes listed, assessors extracted the results of the GRADE approach.

GRADE approach

The GRADE approach has been used in recent Cochrane reviews to rate the certainty of the evidence for an individual outcome.[18] The certainty of evidence reflects the extent of our confidence that the estimates of the effect are correct (or that the true effect lies within a particular range or on one side of a threshold). According to the GRADE approach, evidence from randomized trials is considered to generate high certainty of evidence. This initial GRADE is modified using five key criteria: risk of bias, indirectness, imprecision, inconsistency, and publication bias. A specific outcome is graded at a final rating of "high", "moderate", "low", or "very low" certainty of evidence depending on whether these criteria are achieved.[18,19]

In this context, high certainty of evidence means we are very confident that the true effect lies close to the estimate. Moderate certainty suggests we are moderately confident in the effect estimate, but there is a possibility that it is substantially different. Low certainty means our confidence in the estimate is limited, implying the true effect may differ substantially from the estimate. Very low certainty indicates we have very little confidence in the estimate, suggesting that the true effect is likely to be substantially different.[19,20]

Two independent assessors performed the GRADE approach for the primary outcomes. If the review had already presented GRADE results, these results were used. In cases of disagreement, a third assessor was consulted. The certainty of evidence was downgraded by one level for each of the following criteria:

- Risk of bias: We downgraded the certainty of evidence by one level when 25 % of the participants were from studies judged as having high risk of bias (i.e., one of the following criteria judged as having "high" or "unclear" risk of bias: random allocation, allocation concealment, blinding procedures, and adequate follow-up).[21,22]

- Inconsistency: We downgraded the certainty of evidence by one level when the heterogeneity of pooled estimates was higher than moderate ($I^2 > 40$ %) or ≤ 75 % of participants from studies with findings in the same direction.[23]

- Indirectness: We downgraded the certainty of evidence by one level when > 25 % of the participants included in the meta-analysis were from studies where representativeness was low, including population (i.e., when participants were considered outside of those defined in the inclusion criteria or were limited to particular participants or settings), intervention (i.e., studies only assessed particular versions of the intervention, e.g., a particular dosing or interventions implemented only by expert, highly trained specialists), comparators (i.e., comparisons that were not highly applicable or comparators that were less effective than standard treatment in most settings), and outcome measures (i.e., outcomes that were not the most informative way of measuring effects of the interventions, e.g. using surrogate outcomes or reporting only endpoints).[24]

- Imprecision: We downgraded the certainty of evidence by one level when the total of events was lower than 300 for dichotomous data and the total number of participants was lower than 400 for continuous data.[20]

- Publication bias: Assessment was performed only if funnel plots were available. The certainty of the evidence was downgraded by one level when visual inspection of the funnel plots suggested publication bias.[24]

When a review presented only one randomized study (with < 300 participants), we considered it inconsistent and imprecise, and provided low certainty of evidence. This was further downgraded to very low certainty of evidence if there were also limitations in the design of the trial.

Conclusiveness

A Cochrane review was considered conclusive if it presented high certainty of evidence for at least one primary outcome. Otherwise, if the certainty of evidence was rated as moderate, low, or very low, the review was classified as inconclusive. When a review reported more than one primary outcome, GRADE assessment for the first three primary outcomes were analyzed and the highest certainty of evidence was used. When a review reported more than one comparator intervention, we prioritized them in the following order: placebo intervention, waiting list, minimal intervention (e.g., education), usual care, and an alternative intervention.

Data analysis

We computed the proportion of reviews classified in each category of GRADE certainty of evidence and classified the reviews as conclusive or inconclusive across all four periods (i.e. 2001–2005, 2006–2010, 2011–2015 and 2016–2020). Analysis of all extracted data was completed using Microsoft Excel 2007 (Microsoft Corporation, Redmond, Washington). Results are expressed as median (interquartile range) or frequency (percentage).

Results

Characteristics of the included reviews are provided in Table 1. Overall, the most frequent subdisciplines were cardiothoracic ($n = 33$, 16.5 %), neurology ($n = 33$, 16.5 %), musculoskeletal ($n = 28$, 14 %). and "other" ($n = 36$, 18 %). Reviews categorized as "other" consisted of subdisciplines that did not fall into those listed in the PEDro database,

Table 1

Descriptive characteristics of included reviews ($n = 200$).

Extracted data	2001–2005	2006–2010	2011–2015	2016–2020
Subdiscipline, n (%)				
Cardiothoracic	8 (16)	7 (14)	11 (22)	7 (14)
Continence/ women's health	4 (8)	3 (6)	6 (12)	4 (8)
Ergonomic/ occupational health	0 (0)	1 (2)	1 (2)	2 (4)
Geriatrics	2 (4)	3 (6)	3 (6)	1 (2)
Musculoskeletal	8 (16)	4 (8)	7 (14)	9 (18)
Neurology	5 (10)	16 (32)	6 (12)	6 (12)
Oncology	0 (0)	2 (4)	0 (0)	3 (6)
Orthopedics	6 (12)	3 (6)	6 (12)	1 (2)
Pediatrics	9 (18)	2 (4)	2 (4)	5 (10)
Sports	0 (0)	0 (0)	0 (0)	1 (2)
Other	8 (16)	9 (18)	8 (16)	11 (22)
Continent, n (%)				
Asia	2 (4)	1 (2)	6 (12)	3 (6)
Europe	24 (48)	37 (74)	26 (52)	26 (52)
North America	16 (32)	5 (10)	6 (12)	4 (8)
Oceania	8 (16)	6 (12)	8 (16)	14 (28)
South America	0 (0)	1 (2)	4 (8)	3 (6)
Participants, n (%)				
Acute condition(s)	11 (22)	15 (30)	16 (32)	12 (24)
Chronic condition (s)	31 (62)	25 (50)	25 (50)	32 (64)
Healthy adults (age $\geq$25y)	1 (2)	0 (0)	0 (0)	0 (0)
Other	7 (14)	10 (20)	9 (18)	6 (12)
Intervention, n (%)				
Conservative management	2 (4)	1 (2)	2 (4)	0 (0)
Device	9 (18)	9 (18)	12 (24)	7 (14)
Education	4 (8)	3 (6)	3 (6)	1 (2)
Exercise	5 (10)	11 (22)	10 (20)	14 (28)
Hospital intervention	1 (2)	1 (2)	1 (2)	2 (4)
Mixed intervention	3 (6)	6 (12)	3 (6)	7 (14)
Multiple interventions	4 (8)	7 (14)	3 (6)	8 (16)
Rehabilitation therapy	17 (34)	9 (18)	12 (24)	8 (16)
Self-management	1 (2)	1 (2)	1 (2)	1 (2)
Other	4 (8)	2 (4)	3 (6)	2 (4)
Comparators, n (%)				
Alternative intervention(s)	12 (24)	11 (22)	8 (16)	12 (24)
Multiple comparators	20 (40)	26 (52)	29 (58)	31 (62)
No intervention	6 (12)	7 (14)	7 (14)	3 (6)
Placebo	1 (2)	0 (0)	0 (0)	0 (0)
Standard treatment	4 (8)	2 (4)	2 (4)	3 (6)
Usual care	7 (14)	4 (8)	4 (8)	1 (2)
Identified primary outcomes, n (%)	25 (50)	37 (74)	47 (94)	50 (100)
Number of primary outcomes, median [IQR]	5 [2 to 7]	3 [1 to 5]	3 [2 to 5]	2 [2 to 4]
Number of trials, median [IQR]	8 [5 to 18]	11 [4 to 23]	10 [5 to 18]	14 [8 to 28]
Number of participants, median [IQR]	561 [222 to 2662]	1085 [225 to 2382]	836 [186 to 1673]	1389 [427 to 2397]

IQR = interquartile range.

such as gastrointestinal or vascular conditions. Europe was the most frequent continent of origin, representing 56.5 % ($n = 113$) of reviews in the sample. 113 reviews (56.5 %) examined participants with chronic conditions, followed by 54 (27 %) reviews examining participants with acute conditions. Notably, only one study (< 1 %) examined explicitly healthy individuals. Rehabilitation therapies ($n = 46$, 23 %), exercise ($n = 40$, 20 %), and medical devices ($n = 37$, 18.5 %) were the most common interventions, and study interventions were most often examined against multiple comparators ($n = 106$, 53 %), an alternative

intervention (n = 43, 21.5 %). or no intervention (n = 23, 11.5 %). The number of reviews that clearly specified primary outcomes composed 79.5 % (n = 159) of the sample. This number increased over time, as all reviews from 2016 to 2020 (n = 50) specified primary outcomes compared to half of studies from 2001 to 2005 (n = 25). Alongside this increase, the number of primary outcomes per review trended downwards, with a median of five (IQR 2 to 7) in 2001–2005 to only two (IQR 2 to 4) in 2016–2020.

The percentages of primary outcomes with each GRADE rating by comparison are shown in Fig. 1. Overall, 480 primary outcomes were evaluated, as reviews often presented more than one primary outcome. Low certainty of evidence was the most frequent level of GRADE and composed 55 % (n = 262) of all outcomes. This was followed by moderate (22 %, n = 106), very low (21 %, n = 102), and high certainty (2 %, n = 10). Outcomes with very low certainty of evidence appeared to increase by 20 % over the 20-year period, from 14 % (n = 17) in 2001–2005 to 34 % (n = 40) in 2016–2020. Low-certainty outcomes remained consistent in proportion over time, fluctuating between 38 % and 62 % (n = 45 to 77) per period. Moderate-certainty outcomes also remained unchanged over time and ranged from 15 % to 27 % (n = 18 to 31). The percentage of high-certainty outcomes was also consistent, and its portion was never larger than 4 % (n = 5) per period.

The category and frequency of downgrades for each outcome according to the GRADE criteria are described in Table 2. Overall, 875 downgrades were given, with a mean of 1.8 downgrades per outcome. Risk of bias was the most common reason for an outcome downgrade, representing 45 % (n = 394 of all downgrades), followed by imprecision (38 %, n = 335), and inconsistency (11 %, n = 100). The number of downgrades given for indirectness appeared to increase by 8 %, from one (< 1 %) in 2001–2005 to 20 (9 %) in 2016–2020. The remaining downgrade criteria did not exhibit noticeable changes over time.

Fig. 2 describes the percentage of conclusive reviews (A) and the reporting of need for further studies (B). Total percentage of conclusive studies represented 3 % (n = 6) of the entire sample. Only one review was deemed conclusive during each of the first three time periods (% per period), eventually rising to three reviews (6%) in 2016–2020. Reviews reporting no need for further studies also remained few, with a range of three to seven studies (6 to 14%) per period. Regarding the complete sample (i.e., n = 200), studies that reported no further need composed 9% (n = 18) overall. Neither conclusiveness nor need for further studies appeared to change over time.

Table 2
Frequency of outcome downgrades by criterion, n (%).

| Criteria	2001–2005 n = 220	2006–2010 n = 208	2011–2015 n = 235	2016–2020 n = 212
Risk of bias	109 (50)	99 (48)	101 (43)	85 (40)
Inconsistency	20 (9)	27 (13)	28 (12)	25 (12)
Indirectness	1 (< 1)	5 (2)	10 (4)	20 (9)
Imprecision	90 (41)	74 (36)	94 (40)	77 (36)
Publication bias	0 (0)	3 (1)	2 (1)	5 (2)

Note: Up to three primary outcomes were evaluated per review.

Despite the low percentage of conclusive reviews, there has been accumulation in the absolute number of conclusive reviews across the 20-year period, since a conclusive review is unlikely to later become inconclusive. Given that we sampled 200 reviews, we estimated the accumulation rate of conclusive Cochrane reviews based on the total number of published reviews available in PEDro per year (Fig. 3).

Discussion

Our findings indicate that conclusive Cochrane reviews in physical therapy comprised a very small percentage of the sample, which did not change from 2001 to 2020. Additionally, the reporting of a need for further studies also remained consistent throughout the period, with most reviews indicating further research is needed to support a definitive recommendation.

One potential reason for the percentage of conclusiveness being small may be the large number of outcomes downgraded for risk of bias. The high prevalence of this downgrade criterion has been shown in other studies that evaluated risk of bias of systematic reviews in physical therapy.[6,9,25] Compared to other fields of medicine, addressing risk of bias in randomized trials in the physical therapy field can be challenging,[25] particularly when considering limitations for blinding and development of sham therapies.[9,25] According to GRADE, conclusiveness will be limited when risk of bias is present, as a risk of bias limits the certainty of the evidence. Because achieving the risk of bias criterion leads to a downgrade, this means that in some cases, an outcome with moderate certainty of evidence may lend itself more confidence when acknowledging the inability of the intervention or therapist to be blinded. This acknowledgement is particularly applicable to reviews
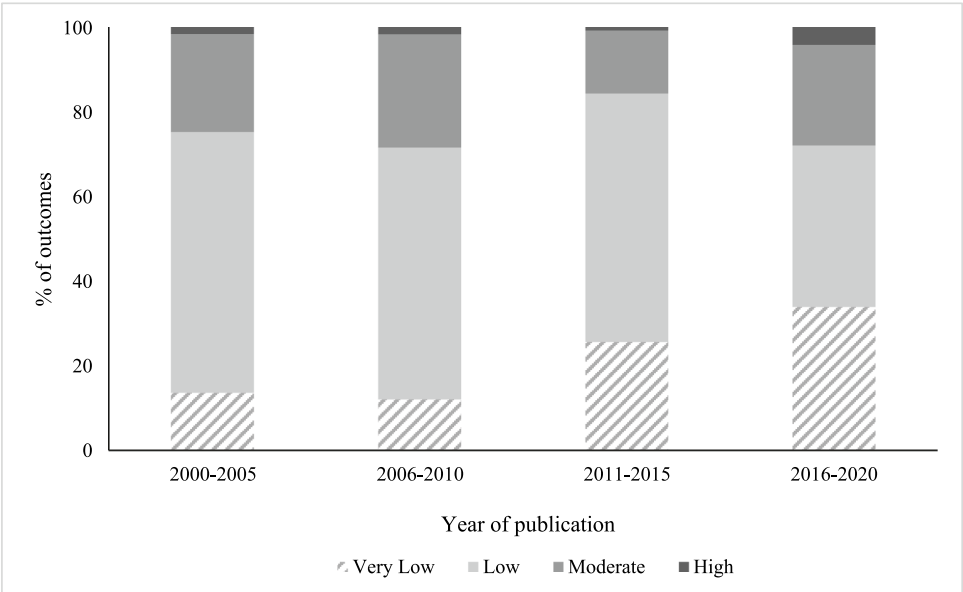
Fig. 1. GRADE distribution for primary outcomes reported from reviews in each five-year period.
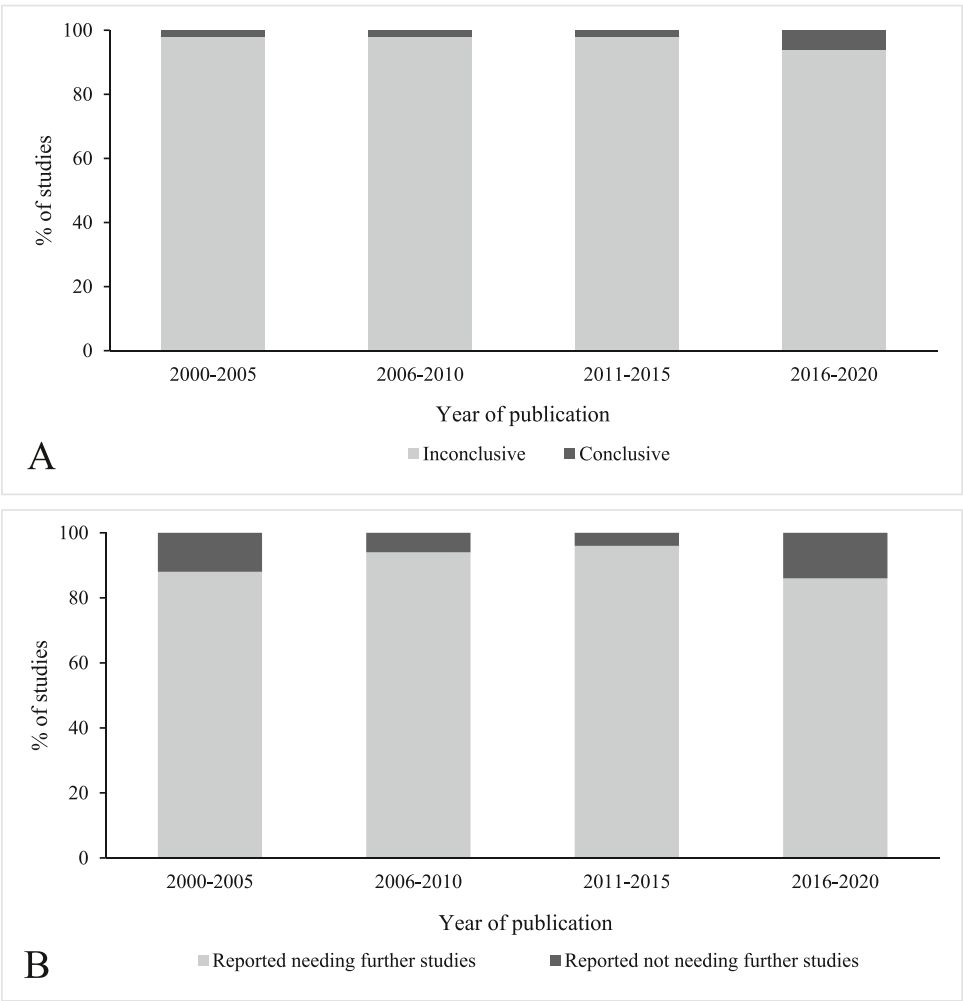
Fig. 2. Percentage of reviews that (A) were deemed conclusive or not, and (B) reported needing further studies or not.
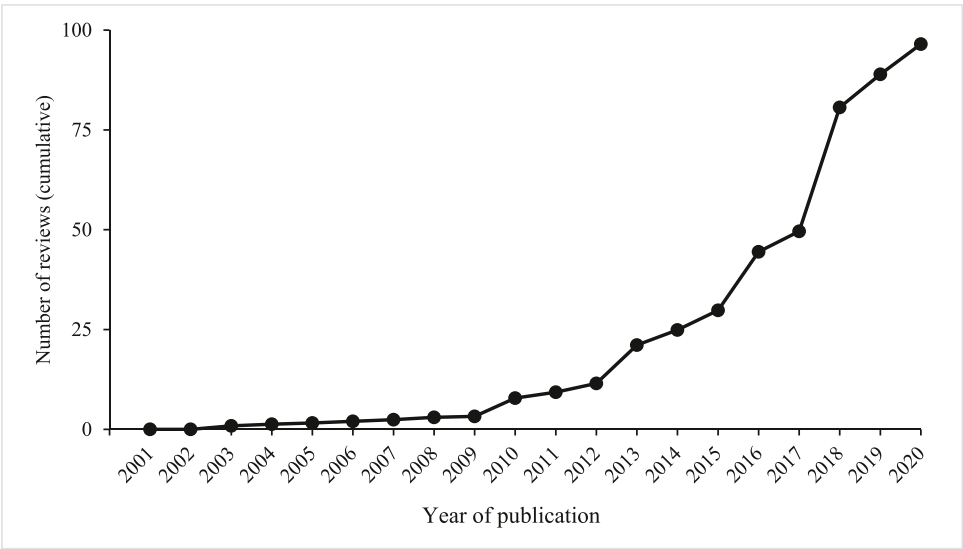

Fig. 3. Extrapolated estimate of the cumulative number of conclusive Cochrane reviews on PEDro published from 2001 to 2020.

examining, for example, manual therapy or exercise as an intervention. Exercise was the second-most common intervention used in our sample therapies and is one of the cornerstones of physical therapy practice.

Our GRADE results lead to a larger discussion for researchers and clinicians in the physical therapy field,[26] that it is important not to disregard exercise outcomes that appear to have moderate certainty on systematic reviews, as the downgrade may only be attributed to risk of bias, because exercise interventions will always face an obstacle to

blinding. It is important that clinicians pay attention to the reasoning behind each given GRADE and interpret the results of the systematic review in this context.

One potential reason for the percentage of conclusiveness not improving may be the steady rise in new review topics being introduced by the Cochrane collaboration each year. In the physical therapy field, approximately 40 new topics are added each year. Increasing the number of physical therapy Cochrane reviews might naturally decrease the proportion of those deemed conclusive. However, because a conclusive review is unlikely to later become inconclusive, it is assumed that conclusive reviews will continue to accumulate.

Another reason for high inconclusiveness may also be the increase in indirectness downgrades. A study was considered indirect if it demonstrated 25% or more heterogeneity in one or more study design areas. The rise in indirectness downgrades may be attributed to the increase in available data overall, as more studies are conducted over time. While not every study will focus on the same population or intervention, researchers may have pooled these together in their analyses, contributing to inconsistencies in the findings and therefore a downgrade in certainty of evidence. For instance, 53% of reviews in our sample analyzed interventions against multiple comparators, creating heterogeneity in the results. Consistent with our findings, a previous study that evaluated systematic reviews in musculoskeletal physical therapy found over half of included reviews contained trials that failed to meet the external validity standard according to the PEDro scale.[27] Because indirectness symbolizes a lack of specificity and applicability, this creates problems in forming conclusions when results are derived from overlapping populations or interventions.

The increase in indirectness downgrades may also be attributed to improved competency in assessing the less commonly achieved criteria (e.g. indirectness, inconsistency, publication bias). Guidelines published in the *Journal of Clinical Epidemiology*[20,28] have provided greater clarity on correctly identifying the presence of each GRADE criterion. The increased awareness may have improved assessors' recognition, leading to greater downgrades given for less obvious criteria, such as indirectness.

Study limitations

Some limitations of this study are that we only extracted Cochrane reviews that were available in PEDro, as they were most relevant to the physical therapy field compared to reviews in other databases. Therefore, our findings might not be applicable to non-Cochrane reviews. We also opted to draw a random sample from each of the four time periods. For future studies, it is suggested that reviews with subsequent updates may be analyzed, assessing the conclusiveness of evidence of the same review across time.

Conclusion

Conclusiveness of Cochrane systematic reviews in physical therapy is low, and the percentage of conclusive reviews has not changed over time, although conclusive reviews are accumulating. Only 3% of studies were deemed conclusive based on the certainty of evidence and 91% of studies reported a need for further research. The findings of this study are important for both researchers and clinicians. Further efforts to improve physical therapy clinical trials are needed to improve conclusiveness.

Ethical approval

N/A

Funding

Declaration of competing interest

Authors declare no competing interest.

References

1. Cumpston M, Li T, Page MJ, et al. Updated guidance for trusted systematic reviews: a new edition of the cochrane handbook for systematic reviews of interventions. *Cochrane Database Syst Rev*. 2019;(10). https://doi.org/10.1002/14651858. ED000142.
2. Moseley AM, Elkins MR, Herbert RD, Maher CG, Sherrington C. Cochrane reviews used more rigorous methods than non-Cochrane reviews: survey of systematic reviews in physiotherapy. *J Clin Epidemiol*. Oct 2009;62(10):1021–1030. https://doi.org/10.1016/j.jclinepi.2008.09.018.
3. Useem J, Brennan A, LaValley M, et al. Systematic differences between cochrane and non-cochrane meta-analyses on the same topic: a matched pair analysis. *PloS one*. 2015;10(12), e0144980. https://doi.org/10.1371/journal.pone.0144980.
4. Tricco AC, Tetzlaff J, Pham B, Brehaut J, Moher D. Non-Cochrane vs. Cochrane reviews were twice as likely to have positive conclusion statements: cross-sectional study. *J Clin Epidemiol*. Apr 2009;62(4):380–386. https://doi.org/10.1016/j. jclinepi.2008.08.008. e1.
5. Mandel D, Littner Y, Mimouni FB, Lubetzky R. Conclusiveness of the cochrane neonatal reviews: a systematic analysis. *Acta Paediatr (Oslo, Norway: 1992)*. Oct 2006;95(10):1209–1212. https://doi.org/10.1080/08035250600580537.
6. Momosaki R, Tsuboi M, Yasufuku Y, et al. Conclusiveness of Cochrane Reviews in physiotherapy: a systematic search and analytical review. *Int J Rehabil Res*. Jun 2019;42(2):97–105. https://doi.org/10.1097/mrr.0000000000000338.
7. Zhang X, Wu Z, Zhao H, Li C, Wu J, Dai G. Conclusiveness of the cochrane reviews in palliative and supportive care for cancer. *Am J Hosp Palliat Care*. Feb 2017;34(1): 53–56. https://doi.org/10.1177/1049909115605058.
8. Cohen S, Lubetzky R, Mimouni FB, Marom R, Mandel D. Conclusiveness of the cochrane reviews in pediatric-gastroenterology: a systematic analysis. *Eur J Gastroenterol Hepatol*. Feb 2013;25(2):252–254. https://doi.org/10.1097/MEG.0b013e32835a1083.
9. Fregni F, Imamura M, Chien HF, et al. Challenges and recommendations for placebo controls in randomized trials in physical and rehabilitation medicine: a report of the international placebo symposium working group. *Am J Phys Med Rehabil*. Feb 2010; 89(2):160–172. https://doi.org/10.1097/PHM.0b013e3181bc0bbd.
10. Moseley AM, Herbert RD, Maher CG, Sherrington C, Elkins MR. Reported quality of randomized controlled trials of physiotherapy interventions has improved over time. *J Clin Epidemiol*. Jun 2011;64(6):594–601. https://doi.org/10.1016/j. jclinepi.2010.08.009.
11. Scurlock-Evans L, Upton P, Upton D. Evidence-based practice in physiotherapy: a systematic review of barriers, enablers and interventions. *Physiotherapy*. Sep 2014; 100(3):208–219. https://doi.org/10.1016/j.physio.2014.03.001.
12. da Silva TM, Costa Lda C, Garcia AN, Costa LO. What do physical therapists think about evidence-based practice? A systematic review. *Man Ther*. Jun 2015;20(3): 388–401. https://doi.org/10.1016/j.math.2014.10.009.
13. Villas Boas PJ, Spagnuolo RS, Kamegasawa A, et al. Systematic reviews showed insufficient evidence for clinical practice in 2004: what about in 2011? The next appeal for the evidence-based medicine age. *J Eval Clin Pract*. Aug 2013;19(4): 633–637. https://doi.org/10.1111/j.1365-2753.2012.01877.x.
14. Hultcrantz M, Rind D, Akl EA, et al. The GRADE Working Group clarifies the construct of certainty of evidence. *J Clin Epidemiol*. Jul 2017;87:4–13. https://doi. org/10.1016/j.jclinepi.2017.05.006.
15. Freire A, Elkins MR, Ramos EMC, Moseley AM. Use of 95% confidence intervals in the reporting of between-group differences in randomized controlled trials: analysis of a representative sample of 200 physical therapy trials. *Braz J Phys Ther*. Jul-Aug 2019;23(4):302–310. https://doi.org/10.1016/j.bjpt.2018.10.004.
16. Yamato TP, Maher CG, Saragiotto BT, Hoffmann TC, Moseley AM. How completely are physiotherapy interventions described in reports of randomised trials? *Physiotherapy*. Jun 2016;102(2):121–126. https://doi.org/10.1016/j. physio.2016.03.001.
17. Elkins MR. Assessing baseline comparability in randomised trials. *J Physiother*. Oct 2015;61(4):228–230. https://doi.org/10.1016/j.jphys.2015.07.005.
18. Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*. 2008;336(7650): 924–926. https://doi.org/10.1136/bmj.39489.470347.AD.
19. Guyatt GH, Oxman AD, Kunz R, Vist GE, Falck-Ytter Y, Schünemann HJ. What is "quality of evidence" and why is it important to clinicians? *BMJ*. 2008;336(7651): 995–998. https://doi.org/10.1136/bmj.39490.551019.BE.
20. Guyatt GH, Oxman AD, Kunz R, et al. GRADE guidelines 6. Rating the quality of evidence–imprecision. *J Clin Epidemiol*. Dec 2011;64(12):1283–1293. https://doi. org/10.1016/j.jclinepi.2011.01.012.
21. Oliveira CB, Maher CG, Ferreira ML, et al. Epidural corticosteroid injections for lumbosacral radicular pain. *Cochrane Database Syst Rev*. Apr 9 2020;4(4), Cd013577. https://doi.org/10.1002/14651858.Cd013577.
22. Henschke N, Ostelo RW, van Tulder MW, et al. Behavioural treatment for chronic low-back pain. *Cochrane Database Syst Rev*. Jul 7 2010;2010(7), Cd002014. https:// doi.org/10.1002/14651858.CD002014.pub3.
23. Higgins JP Gse. Cochrane handbook for systematic reviews of interventions version 5.1.0 (updated March). *The Cochrane Collaboration, 2024. handbook.cochrane.org.*. 2024.

24. Ryan RHS. *How to GRADE the quality of the evidence*; 2011. http://cccrg.cochrane.org/author-resources.

25. Hernando DF, Elkins M, Freire A. Reporting of confidence intervals, achievement of intended sample size, and adjustment for multiple primary outcomes in randomised trials of physical therapy interventions: an analysis of 100 representatively sampled trials. *Braz J Phys Ther*. May-Jun 2024;28(3), 101079. https://doi.org/10.1016/j.bjpt.2024.101079.

26. Souto-Miranda S, Brazete Cruz E, Pires D, Ribeiro F, Cordeiro N, Jácome C. Priorities in physical therapy research: a scoping review. *Braz J Phys Ther*. Nov-Dec 2024;28(6), 101135. https://doi.org/10.1016/j.bjpt.2024.101135.

27. 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*. Aug 2021;29(4):203–215. https://doi.org/10.1080/10669817.2020.1839728.

28. Guyatt GH, Oxman AD, Schünemann HJ, Tugwell P, Knottnerus A. GRADE guidelines: a new series of articles in the journal of clinical epidemiology. *J Clin Epidemiol*. Apr 2011;64(4):380–382. https://doi.org/10.1016/j.jclinepi.2010.09.011.