Received: 21 July 2020 | Accepted: 20 August 2020

DOI: 10.1002/tsm2.196

ORIGINAL ARTICLE

WILEY

Risk of bias in systematic reviews of tendinopathy management: Are we comparing apples with oranges?

Dimitris Challoumas | **Neal L. Millar**

Institute of Infection, Immunity and Inflammation, College of Medicine, Veterinary and Life Sciences, University of Glasgow, Glasgow, UK

Correspondence
Neal L. Millar, Institute of Infection, Immunity and Inflammation, College of Medicine, Veterinary and Life Sciences, University of Glasgow, Glasgow G12 8TA, UK.
Email: neal.millar@glasgow.ac.uk

Funding information
This work was funded by grants from the Medical Research Council UK (MR/R020515/1).

We aimed to provide an overview of the use of risk of bias (RoB) assessment tools in systematic reviews (SRs) in tendinopathy management given increased scrutiny of the SR literature in clinical decision making. A search was conducted in Medline from inception to June 2020 for all SRs of randomized controlled trials (RCTs) assessing the effectiveness of any intervention(s) on any location(s) of tendinopathy. Included SRs had to use one of (a) Cochrane Collaboration tool, (b) PEDro scale, or (c) revised Cochrane Collaboration tool (RoB 2) for their RoB assessment. A total of 46 SRs were included. Around half of SRs (46%) did not use an RoB assessment in data synthesis, and only 30% used it to grade the certainty of evidence. The RoB 2 tool was the most likely to determine "overall high RoB" (52%) followed by the Cochrane Collaboration tool (34.6%) and the PEDro scale (18.6%) as determined by the authors of the SRs. We have demonstrated substantial problems associated with the use of RoB assessments in tendinopathy SRs. The universal use of a single RoB assessment tool should be promoted by journals and SR guidance documents.

KEYWORDS
musculoskeletal system, tendon injury

1 | INTRODUCTION

The constant emergence of new treatment modalities for tendinopathy over the last few decades and the absence of robust evidence for their effectiveness has led to an increasing number of randomized controlled trials (RCTs). Systematic reviews (SRs) of RCTs constitute the strongest level of evidence and can therefore inform clinical practice, both at a policy level and an individual physician level. A SR should be transparent and reproducible, and subjectivity should be kept to a minimum.[1] Unfortunately, firm guidance on conducting a SR does not exist and several parameters are left to the judgment of the authors. Moreover, recent debate in the Lancet argues that the findings of SRs may be flawed as they often include poor-quality studies that should have not been published in the first place.[2]

One of these parameters is risk of bias (RoB) assessment; not only is it a subjective process in its nature, but the existence of several RoB assessment tools further decreases reproducibility by introducing inconsistency. RoB assessment plays an integral role in SRs, and it is an essential part of data synthesis and the reporting of the results. It can be used in one of two ways in a SR, either for subgroup analyses (ie, including only RCTs with low risk of bias) or in determining the strength of evidence for each result in conjunction with other limitations of the included evidence that arise as a result of combining the findings of different studies (consistency, imprecision, etc).[3]

The Cochrane Collaboration tool[4] for assessing internal validity (RoB), which was introduced in 2008 and is the tool most frequently used in SRs of RCTs, consists of 7 components/questions, which can be rated as "low" risk, "unclear" risk, or "high" risk of bias. Through its use over the last decade, it has been associated with a lot of confusion, low inter-rater reliability, and wrong implementation in SRs.[5] Additionally, the creators did not specify how the tool should be used to determine overall RoB for each assessed RCT and instead they advised an overall judgment of the result at a domain and not study level, which is both impractical and very subjective. The second most commonly used tool, the PEDro scale,[6,7] is a scoring system that can be used to determine overall RoB for each study based on the overall score out of 10. It includes all the domains of the Cochrane tool and some additional items, and unlike the Cochrane tool, it is less subjective as the assessor only has two possible answers for each item/question: "yes" or "no." The main disadvantage of its simplicity, however, is that methodological aspects of the assessed RCT that are not described clearly in the article are automatically scored with a "no," whereas the Cochrane tool has an "unclear" option, which again is not clear how it should be used in the determination of the overall RoB.

The Cochrane group has recently published a revised RoB assessment tool, the RoB 2,[5] which, according to the authors, is less subjective, more reproducible, and has more direct implementations in data synthesis. It is made up of 5 items/questions, and each one has a number of signaling questions, which help the author reach a final conclusion about the RoB in each item according to a pre-defined formula. This can either be "low" risk, "high" risk, or "some concerns." The creators, having realized the importance of determining overall RoB for each study for practical and reproducible implementation of the RoB assessment in data synthesis, have also described how decisions on overall RoB for each study should be reached. Finally, they highlight that RoB should be assessed on an outcome level for each included RCT.[5]

The introduction of the new RoB assessment tool, regardless of whether it is more effective or not than other tools at predicting the actual RoB, is expected to further increase inconsistency across different SRs. This has the potential to lead to conflicting conclusions between SRs assessing and comparing the same interventions with regard to the strength of evidence of the results and can cause confusion in the translation of the findings and their implementation in clinical practice.

The aims of the present were (a) to provide an overview of the use of RoB assessment tools in SRs of RCTs in tendinopathy through a scoping review and (b) to assess inter-tool reliability among the Cochrane Collaboration tool, the revised Cochrane Collaboration tool (RoB 2), and the PEDro scale at determining overall RoB in tendinopathy SRs. Finally, we provide recommendations at an RCT level, SR level, and journal level with an ultimate objective to make RoB assessment and its use in data syntheses as understandable, transparent, objective, and reproducible as possible.

2 | METHODS

2.1 | Eligibility

SRs were eligible if they assessed the effectiveness of any intervention(s) on any location(s) of tendinopathy in patients over 16 years of age, included only RCTs, and used one of the following RoB assessment tools: Cochrane Collaboration tool, PEDro scale, RoB 2 tool (revised Cochrane Collaboration tool). Exclusion criteria included SRs including a mixture of randomized and non-randomized studies and a mixture of participants with tendinopathy and other conditions. SRs in languages other than English were also excluded. No criteria were used regarding the following parameters: publication date, journal type, type of tendinopathy and intervention, outcome measures, and length of follow-up.

2.2 | Search strategy—Screening

A literature search was conducted by the first author via Medline in June 2020 with the following Boolean operators in "All Fields": "((systematic review) OR (meta-analysis) AND (tendin*) AND (randomi*)).

For all eligible articles, the reference lists and PubMed's "similar articles" list were screened to identify potentially eligible articles that may have been missed at the initial search. Figure 1 (PRISMA flowchart) illustrates the article screening process.

The initial search returned a total of 208 articles. After exclusion of non-eligible articles according to our pre-defined criteria and inclusion of articles identified from reference screening, 46 SRs were included in our review.

2.3 | Data Extraction—Handling

2.3.1 | Scoping review

The included SRs were read by the first author, and data were extracted in a Microsoft Word table regarding the following: (a) general SR characteristics (number of included RCTs, location(s) of tendinopathy, intervention(s) assessed, key findings), (b) RoB assessment tool used, (c) whether an overall RoB was determined for each assessed RCT, (d) whether RoB assessment was performed on a study or outcome level, and (e) how RoB assessment was used in data syntheses.

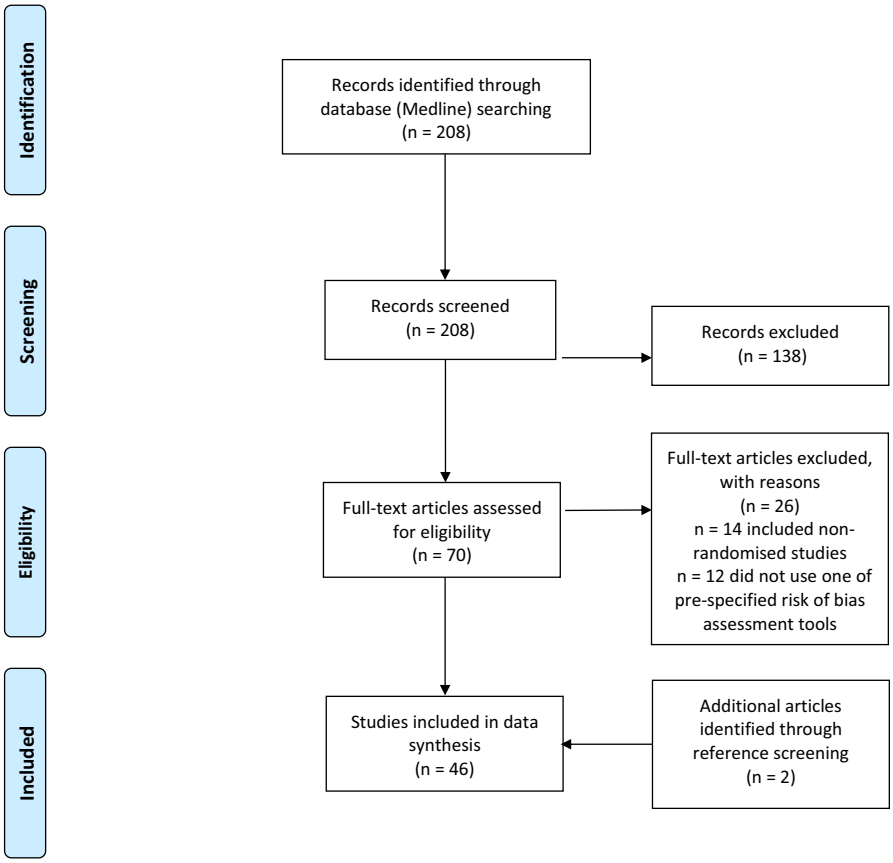

FIGURE 1 PRISMA flow diagram [Colour figure can be viewed at wileyonlinelibrary.com]

2.3.2 | Assessment of consistency of risk of bias assessment

In order to assess for disparity of tools determining overall RoB, we used two separate methods. Firstly, we calculated the proportion of RCTs assessed in all included SRs being determined as of "high overall RoB" for each one of the 3 tools separately and the mean proportion for each tool. Where overall RoB was determined by the authors of the original SR for each RCT, this was used. We also used our own pre-defined criteria (see below) to determine overall RoB for each RCT based on the RoB assessment results reported by the SR authors. Inter-tool reliability was not evaluated formally with statistical tests for this method as the RCTs assessed by each tool were not the same; instead, our purpose was to give a general impression on the likelihood of each tool to determine "high overall RoB" for RCTs and investigate for inter-rater inconsistencies when different criteria are used for the same studies.

Secondly, in light of the newly published RoB 2 tool by the Cochrane Collaboration and its use by the most recently published SR of RCTs in Achilles tendinopathy by van der Vlist et al,[8] we assessed RoB of its 29 included RCTs using the two other RoB assessment tools, the Cochrane Collaboration tool and the PEDro scale. We then compared the reliability among the three tools (Cochrane Collaboration and PEDro as performed by the authors of the present review and RoB 2 by the authors of the original SR) at determining overall RoB. We only tested inter-tool reliability for overall RoB determination and not specific domains of the tools as only the former is directly associated with implementation of RoB assessment in data synthesis.

Inter-tool reliability was only assessed for determining "high overall RoB," which is the aspect of RoB assessment with direct application in data syntheses. "High overall RoB" RCTs determine downgrading of the quality of the evidence, and they are the studies removed for subgroup/sensitivity analyses. For the purposes of the statistical tests, the 29 assessed RCTs were divided in two categories, "high overall RoB" and "other" ("low overall RoB"/"unclear RoB"/"some concerns"), and each category represented each one of the two possible outcomes in the Cohen's kappa formulas.

Overall RoB determination (our criteria)

The RoB 2 tool provides clear, specific instructions on how the overall RoB for each study should be determined[5]; therefore, we only used the SR authors' assessment.

With regard to the PEDro scale, its final score is traditionally interpreted as 8-10 "excellent quality" and 6-7 "good quality"; therefore, we used ≥6 as a cutoff to divide high and low overall RoB (or low and high study quality, respectively) firstly as this is the criterion most commonly used by SR authors (PEDro ≥ 6). We also used ≥8 as a cutoff to

see which score gives more similar results to the other tools (PEDro ≥ 8). As the majority of authors use the PEDro scale for "study quality" and not RoB assessment, for the purposes of this review "high overall RoB" was synonymous to "moderate" or "poor" study quality.

For the Cochrane Collaboration tool, RCTs were considered as "high overall RoB" if they had: (a) high RoB in any of "random sequence generation," "allocation concealment," "blinding of patients and staff," or "blinding of outcome measures" or (b) high RoB in 2 or more of the remaining 3 items ("completeness of outcome data," "selective reporting," and "other") or (c) high RoB in one of the 3 remaining domains if the authors felt the RoB introduced through that domain was significant enough to affect the results of the study. "Unclear overall RoB" was assigned to studies with 3 or more unclear RoB in individual domains not fulfilling the criteria for "high overall RoB," and "low overall RoB" in those not fulfilling the criteria for high and unclear overall RoB. These criteria, especially for the Cochrane tool and to a lesser extent for the PEDro scale, have been specified by the authors of the present review based on advice deriving from the creators of the Cochrane tool and other researchers[9-11]; they do not represent the "appropriate" criteria as the creators themselves did not specify any; however, we use them to emphasize the extent of inconsistency and subjectivity.

2.4 | Statistical analysis

Cohen's kappa statistic was used to assess inter-tool reliability at determining "high overall RoB." According to the value of the statistic (range 0-1), the strength of agreement can be: equivalent to chance (0), slight (0.1-0.2), fair (0.21-0.4), moderate (0.41-0.6), substantial (0.61-0.8), near perfect (0.81-0.99), perfect (1).

The following formula was used for the calculation of Cohen's statistic between each combination of two tools:

$$k = (Po - Pe)/(1 - Pe)$$

where *Po*: the sum of the mutual RCTs rated as "high overall RoB" and "other" in the two tools; *Pe*: (proportion of "high overall RoB" RCTs multiplied by proportion of "other" RCTs in tool 1) + (proportion of "high overall RoB" RCTs multiplied by proportion of "other" RCTs in tool 2).

3 | RESULTS

3.1 | Scoping review

Table 1 summarizes the key characteristics of the eligible SRs.[8,12-56] Of the 46 included SRs, 31 used the Cochrane Collaboration tool, 13 the PEDro scale, 2 the revised Cochrane Collaboration tool (RoB 2), and 2 both the Cochrane Collaboration tool and the PEDro scale. Modified versions of the PEDro scale and the Cochrane Collaboration tool were used by two and one SRs, respectively. RoB was assessed on an outcome and not study level in only 3 SRs (6.5%). An overall RoB for each assessed RCT/outcome was determined in 17 SRs (37%; n = 7 PEDro scale, n = 2 RoB 2 tool, n = 8 Cochrane Collaboration tool). A total of 21 SRs (46%) did not use the results of their RoB assessment anywhere in data synthesis; the remaining 25 that did used it for either subgroup/sensitivity analyses excluding "high overall RoB"/"low-quality" studies (n = 9; 36%), for grading the quality of the evidence (n = 14; 56%), or both (n = 1; 4%). Where the quality of the evidence was graded, tools used included the GRADE tool[3] (n = 6; 43%), the Cochrane BRG tool[9] (n = 5; 36%), and the NHMRC tool[1] (n = 1; 7%), while the authors of 3 SRs (21%) graded the evidence arbitrarily without a pre-specified method.

3.1.1 | Overall RoB determination

Where authors of SRs determined overall RoB of assessed RCTs, the following methods were used for each tool:

- RoB 2: according to the instructions of the tool (n = 2)
- Cochrane Collaboration tool: (a) "overall high RoB" where <3 domains had low RoB (n = 2) or where >3 domains had high RoB (n = 1); (b) "overall low RoB" where the total score of the study was >70% (out of 16; low RoB scored 2, unclear RoB 1, and high RoB 0, n = 1); (c) "good quality study" where no more than 1 domains of the tool, precision and external validity were high RoB (n = 2); (d) method not described (n = 2)
- PEDro: (a) "overall good quality/low RoB" where total score ≥6/10 (n = 4), ≥7/10 (n = 1 lee) or ≥7/13 for modified PEDro (n = 1); (b) "overall low quality/high RoB" where total score < 5/10 (n = 2)

3.2 | Assessment of consistency of risk of bias assessment

Table 2 shows the proportion of "overall high RoB" RCTs as determined by (a) the authors of the original SRs where performed, using their own "high overall RoB" criteria and (b) the first author of the present review (DC) based on the RoB assessment performed by the SR authors using our predefined "high overall RoB" criteria for each tool. Mean percentages were calculated for each tool.

TABLE 1 Key characteristics of included systematic reviews and details on use of risk of bias

Authors	Tendinopathy	Number of included studies	Intervention Assessed	Summary of Findings	RoB Assessment Tool	RoB Assessment on study or outcome level	Method for determining overall RoB	Use of RoB in data synthesis
Arirachakaran et al (2016)	Lateral Elbow	10	PRP, Autologous blood, corticosteroid injection	PRP can improve pain and has fewer complications. Autologous blood can improve pain, function, and pain pressure thresholds but has higher complication rates.	Cochrane	Study	Overall RoB not determined	None
Arirachakaran et al (2017)	Shoulder calcific	7	ESWT, US-guided lavage, corticosteroid injection, and combined treatment	US-guided lavage is the treatment of choice	Cochrane	Study	Overall RoB not determined	None
Bannuru et al (2014)	Shoulder calcific	28	ESWT	High-energy ESWT is effective at improving pain and function	Cochrane	Study	Overall RoB not determined	Subgroup analysis "including high-quality studies"
Bjordal et al (2008)	Lateral Elbow	18	Laser therapy	Laser therapy administered with optimal doses can provide short-term pain relief and improve disability	PEDro	Study	"Good quality" ≥6	Subgroup analysis "excluding low-quality studies"
Boudreault et al (2014)	Shoulder	12	Oral NSAIDs	Oral NSAIDs effective at reducing short-term pain but not function	Cochrane	Study	"Good quality" >70% (scoring system used)	Evidence grading (arbitrary)
Catapano et al (2020)	Shoulder	5	Dextrose Prolotherapy	Prolotherapy is potentially useful adjunct to physical therapy	Cochrane	Study	Overall RoB not determined	None
Challoumas et al (2019a)	All	12	Surgery	Surgery superior to no treatment/placebo but not sham surgery or physiotherapy	Cochrane	Study	Combined assessment of overall RoB, external validity, and precision	Evidence grading (Cochrane BRG)
Challoumas et al (2019b)	All	10	Topical GTN	Topical GTN superior to placebo in medium term	Cochrane	Study	Combined assessment of overall RoB, external validity, and precision	Evidence grading (Cochrane BRG)
Chen et al (2019)	Patellar	11	Non-surgical treatments	LR-PRP is most effective non-surgical treatment	PEDro	Study	Overall RoB not determined	None

(Continues)

TABLE 1 (Continued)

Authors	Tendinopathy	Number of included studies	Intervention Assessed	Summary of Findings	RoB Assessment Tool	RoB Assessment on study or outcome level	Method for determining overall RoB	Use of RoB in data synthesis
Coombes et al (2010)	All	41	Corticosteroid and other injections	Corticosteroid injections are effective in the short-term, other injections may provide long-term benefit for lateral elbow tendinopathy	Modified PEDro	Study	"Good quality" score >6/13	Only "high-quality studies" included in SR
Dan et al (2019)	Patellar	2	Surgery	Inconclusive due to low quality of evidence; surgery likely no more effective than eccentric exercise	Cochrane	Outcome	Overall RoB not determined	Evidence grading (GRADE)
de Vos et al (2014)	Lateral Elbow	6	PRP	PRP not effective	PEDro	Study	"Good quality" ≥6	Evidence grading (Cochrane BRG)
Desjardins-Charbonneau et al (2015a)	Shoulder	10	Taping	Inconclusive due to low quality of evidence	Cochrane	Study	Overall RoB not determined*	None
Desjardins-Charbonneau et al (2015b)	Shoulder	21	Manual therapy	Manual therapy may decrease pain, but it is unclear if it improves function	Cochrane	Study	Overall RoB not determined*	None
Desmeules et al (2016a)	Shoulder	10	Exercise	Exercise is effective at treating workers and promotes return to work	Cochrane	Study	Overall RoB not determined*	None
Desmeules et al (2016b)	Shoulder	6	TENS	Inconclusive due to low quality of evidence	Cochrane	Study	Overall RoB not determined*	None
Desmeules et al (2015)	Shoulder	11	Therapeutic US	Therapeutic US administered with exercise no more superior than exercise alone. Compared to laser treatment it is less effective at alleviating pain	Cochrane	Study	Overall RoB not determined*	None
Dong et al (2015)	Shoulder	33	All	Exercise-based treatments and acupuncture ideal for early disease. Surgery recommended for long-term disease. Corticosteroid injections and laser treatment discouraged.	Cochrane	Study	"High overall RoB" if <3 "low RoB" domains	Subgroup analysis "excluding low-quality studies"
Dong et al (2016)	Lateral Elbow	27	Injection therapies	Some injection therapies can be effective (eg, BOTOX and PRP) but not corticosteroids. Hyaluronate and prolotherapy need more research.	Cochrane	Study	Method not described	Subgroup analysis "excluding low-quality studies"
Fitzpatrick et al (2017)	All	18	PRP	Good evidence to support single injection of PRP under US guidance	Modified Cochrane	Study	High risk if >3 high-risk domains	Subgroup analysis "excluding high RoB studies"

(Continues)

TABLE 1 (Continued)

Authors	Tendinopathy	Number of included studies	Intervention Assessed	Summary of Findings	RoB Assessment Tool	RoB Assessment on study or outcome level	Method for determining overall RoB	Use of RoB in data synthesis
Haslerud et al (2015)	Shoulder	17	Laser therapy	Laser therapy can offer clinically relevant pain relief and improvement in symptoms alone and in combination with physiotherapy	PEDro	Study	"Low quality" if <5	Evidence grading (arbitrary)
Ioppolo et al (2013)	Shoulder calcific	6	ESWT	ESWT effective in terms of pain, function and resorption of calcific deposits	PEDro	Study	Overall RoB not determined	None
Lafrance et al (2019)	Shoulder calcific	3	US-guided lavage	US-guided lavage is more effective than shockwave therapy or a corticosteroid injection alone	Cochrane	Study	Overall RoB not determined*	None
Lee et al (2011)	Shoulder calcific	9	ESWT	Inconclusive due to low quality of evidence	PEDro	Study	"Low risk" if ≥7	Evidence grading (NHMRC)
Li et al (2019)	Lateral Elbow	7	PRP, corticosteroid injection	Corticosteroid injection superior to PRP in short-term but PRP more effective in long-term	Cochrane	Study	Overall RoB not determined	None
Liao et al (2018)	Lower limb tendinopathies	29	ESWT	ESWT is effective for pain and function	PEDro	Study	"Good or excellent quality" ≥6	None
Lin et al (2020)	Shoulder	5	PRP	PRP may be beneficial for long-term pain	Cochrane	Study	Overall RoB not determined	Subgroup analysis "excluding low-quality studies"
Lin et al (2019)	Shoulder	7	Injection therapies	Corticosteroid effective in short but not long-term, PRP and prolotherapy superior in the long-term	Cochrane	Outcome	Method not described	Subgroup analysis "excluding low-quality studies"
Lin et al (2018)	Lateral Elbow	6	Botulinum toxin injection (BOTOX)	BOTOX injections superior to placebo and as effective as corticosteroid injections (though less effective for short-term pain)	Cochrane	Study	Overall RoB not determined	Evidence grading (arbitrary)
Louwerens et al (2014)*	Shoulder calcific	20	Minimally invasive therapies	High-energy ESWT safe and effective in short- and mid-term	Cochrane	Study	Overall RoB not determined	Evidence grading (GRADE)
Martimbianco et al (2020)	Achilles	4	Laser therapy	Inconclusive due to low quality of evidence	Cochrane	Study	Overall RoB not determined	Subgroup analysis "excluding low-quality studies" and evidence grading (GRADE)

(Continues)

TABLE 1 (Continued)

Authors	Tendinopathy	Number of included studies	Intervention Assessed	Summary of Findings	RoB Assessment Tool	RoB Assessment on study or outcome level	Method for determining overall RoB	Use of RoB in data synthesis
Mendonca et al (2020)	Patellar	9	Conservative treatment	Inconclusive due to low quality of evidence	PEDro	Study	"High risk" <5	Evidence grading (GRADE)
Miller et al (2017)	All	16	PRP	PRP more efficacious than control	Cochrane	Study	Overall RoB not determined	None
Mohamadi et al (2017)	Shoulder	14	Corticosteroid injections	Corticosteroid injections provide minimal transient pain relief in a small number of patients	Cochrane, Jadad	Study	Overall RoB not determined	None for Cochrane tool
Murphy et al (2019)	Achilles	7	Heavy eccentric calf training (HECT)	HECT may be superior to no treatment and traditional physiotherapy but inferior to other exercise interventions	RoB 2	Study	According to tool instructions	Evidence grading (GRADE)
Ortega-Castillo & Medina-Porqueres (2016)	Shoulder & Lateral elbow	12	Eccentric exercise	Eccentric exercise effective for pain and strength but its effectiveness compared to other treatments remains questionable	PEDro	Study	Overall RoB not determined	Evidence grading (Cochrane BRG)
Sussmilch-Leitch et al (2012)	Achilles	19	Physical therapies	Eccentric exercise recommended as first line with or without laser therapy. ESWT may be equally effective	Modified PEDro, Cochrane	Study	"High risk" if <3 "low RoB" domains of Cochrane tool	Subgroup analysis "excluding low-quality studies"
Tsikopoulos et al (2016)	All	5	PRP	PRP provided no more clinical benefit than placebo or dry needling	Cochrane	Study	Overall RoB not determined	None
Toliopoulos et al (2014)	Shoulder	15	Surgery	Surgery no more effective than exercises. Arthroscopic surgery may be superior to open for some outcome measures	Cochrane	Study	Overall RoB not determined*	None
Van der Vlist et al (2020)	Achilles	29	All	No clinically relevant difference among treatments at 3 or 12 mo follow-up	RoB 2	Outcome	According to tool instructions	Evidence grading (GRADE)
Wasielewski & Kotsko (2007)	Lower Limb tendinopathies	11	Eccentric exercise	Eccentric exercise may improve pain and strength	PEDro	Study	Overall RoB not determined	None
Woodley et al (2007)	All	11	Eccentric exercise	Inconclusive due to low quality of evidence	PEDro, Cochrane BRG	Study	"High quality if ≥6	Evidence grading (Cochrane BRG)
Wu et al (2017)	Shoulder calcific	14	Non-operative treatments	US-guided needling and ESWT (radial and high-energy focused) alleviate pain and achieve complete resolution of calcium deposits	Cochrane, PEDro	Study	Overall RoB not determined	None
Xiong et al (2019)	Lateral Elbow	4	ESWT vs Corticosteroid	ESWT may be superior to corticosteroids	Jadad, Cochrane	Study	Overall RoB not determined	None

(Continues)

TABLE 1 (Continued)

Authors	Tendinopathy	Number of included studies	Intervention Assessed	Summary of Findings	RoB Assessment Tool	RoB Assessment on study or outcome level	Method for determining overall RoB	Use of RoB in data synthesis
Yan et al (2019)	Lateral Elbow	5	US therapy and ESWT	ESWT superior to US therapy up to 6 mo for pain and pain-free grip strength	Modified Jadad, Cochrane	Study	Overall RoB not determined	None
Zhang et al (2019)	Shoulder calcific	8	US-guided lavage	US-guided lavage may be superior to ESWT in pain relief and calcification clearance	Cochrane	Study	Overall RoB not determined	None

Abbreviations: BRG, back review group; ESWT, extracorporeal shock wave therapy; GRADE, grading of recommendations, assessment, development and evaluations; GTN, glyceryl trinitrate; LR-PRP, leukocyte-rich platelet-rich plasma; NHMRC, national health and medical research council; NSAIDs, no-steroidal anti-inflammatory drugs; PEDro, physiotherapy evidence database scale; PRP, platelet-rich plasma; RoB, risk of bias; TENS, transcutaneous electrical nerve stimulation; US, ultrasound.

*Scoring system used to calculate mean score of all RCTs but cutoffs for high and low risk not specified

3.2.1 | Consistency among tools

Based on the overall RoB assessments reported by the authors of the original SRs, the RoB 2 tool was the most likely to determine a "high overall RoB" (mean proportion of high RoB RCTs 52%), followed by the Cochrane Collaboration tool (mean proportion 34.6%). The PEDro scale was associated with the lowest mean proportion of "high overall RoB" RCTs (18.6%).

When the pre-defined criteria of the authors of the present review were applied, the PEDro ≥ 8 was associated with the highest proportion of high RoB studies (65.4%), followed by the Cochrane Collaboration tool (55%), and finally the PEDro ≥ 6 (29.2%).

3.2.2 | Consistency when different criteria used (SR authors vs authors of present review)

Where we determined "high overall RoB" using our criteria based on the RoB assessment results of the SR authors, the mean proportion of "high overall RoB" studies was substantially higher compared to that of the SR authors for the Cochrane Collaboration tool (55% vs 34.6%) and for the PEDro ≥ 8 (65.4% vs 18.6%). For the PEDro ≥ 6, the difference was less significant (29.2% vs 18.2%) as the majority of SR authors using the PEDro chose a ≥6 cutoff too. The highest variability for individual SRs between the proportion of studies with "high overall RoB" of the SR authors and ours was observed in the Cochrane tool (eg, 3% vs 73% for Dong et al[29]; 0% vs 72% for Fitzpatrick et al[31]) and the PEDro ≥ 8 (eg, 0% vs 82% for Haslerud et al[32]).

3.2.3 | Inter-tool reliability in example systematic review

Tables 3a and 3b shows the RoB assessment that we performed for the 29 RCTs of the van der Vlist[7] SR using the Cochrane Collaboration tool (Table 3a) and PEDro scale (≥6 and ≥ 8) (Table 3b) with our criteria. Table 3c shows the RoB assessment as performed by van der Vlist et al[7] using the RoB 2 tool and the results of the overall RoB assessment from the other two tools as derived from Tables 3a and 3b, highlighting the generally poor inter-tool reliability. The only comparison that produced substantial reliability ($k = 0.76$) was that between the Cochrane tool and the PEDro ≥ 8. Fair reliability was found for the comparisons between the Cochrane tool and the PEDro ≥ 6 ($k = 0.36$), the Cochrane and the RoB 2 ($k = 0.29$), and the RoB 2 and PEDro ≥ 8 ($k = 0.26$). Finally, inter-tool reliability between the RoB 2 and the PEDro ≥ 6 was only slight ($k = 0.03$).

TABLE 2 Determination of high overall RoB with the 3 tools using the systematic review authors' criteria and our criteria

Tool	SR	SR authors' "high overall RoB"	DC "high overall RoB" Cochrane Collaboration	DC "high overall RoB" PEDro	
PEDro				≥6/10	≥8/10
	Bjordal et al (2008)	1/18 (6%)	-	NA	NA
	Chen et al (2019)	ND	-	2/11 (18%)	4/11 (36%)
	Coombes et al (2010)	23/64 (36%)	-	29/64 (45%)	46/64 (72%)
	de Vos et al (2014)	2/6 (33%)	-	2/6 (33%)*	4/6 (66%)
	Haslerud et al (2015)	0/17 (0%)	-	3/17 (18%)	14/17 (82%)
	Ioppolo et al (2013)	ND	-	NA	NA
	Lee et al (2011)	3/9 (33%)	-	3/9 (33%)*	6/9 (66%)
	Liao et al (2018)	0/29 (0%)	-	0/29 (0%)*	13/29 (45%)
	Mendonca et al (2020)	2/9 (22%)	-	3/9 (33%)	5/9 (56%)
	Ortega-Castillo & Medina-Porqueres (2016)	ND	-	2/12 (17%)	10/12 (83%)
	Wasielewski & Kotsko (2007)	ND	-	5/11 (45%)	9/11 (82%)
	Wu et al (2017)	ND	-	NA	NA
	Mean Proportion	18.6%	-	29.2%	65.4%
Cochrane Collaboration	Arirachakaran et al (2016)	ND	7/10 (70%)	-	-
	Arirachakaran et al (2017)	ND	3/7 (43%)	-	-
	Bannuru et al (2014)	ND	NA	-	-
	Boudreault et al (2014)	ND	7/12 (58%)	-	-
	Catapano et al (2020)	ND	3/6 (50%)	-	-
	Challoumas et al (2019a)	ND	9/12 (75%)	-	-
	Challoumas et al (2019b)	ND	6/10 (60%)	-	-
	Dan et al (2019)	ND	2/2 (100%)	-	-
	Desjardins-Charbonneau et al (2015a)	ND	10/10 (100%)	-	-
	Desjardins-Charbonneau et al (2015b)	16/21 (76%)	20/21 (95%)	-	-
	Desmeules et al (2016a)	8/10 (80%)	10/10 (100%)	-	-
	Desmeules et al (2016b)	ND	6/6 (100%)	-	-
	Desmeules et al (2015)	ND	9/11 (82%)	-	-
	Dong et al (2015)	1/33 (3%)	24/33 (73%)	-	-
	Dong et al (2016)	1/27 (4%)	10/27 (37%)	-	-
	Fitzpatrick et al (2017)	0/18 (0%)	13/18 (72%)	-	-
	Lafrance et al (2019)	2/3 (66%)	2/3 (66%)	-	-
	Li et al (2019)	ND	4/7 (57%)	-	-
	Lin et al (2020)	ND	2/5 (40%)	-	-
	Lin et al (2019)	0/7 (0%)	NA	-	-
	Lin et al (2018)	0/6 (0%)	0/6 (0%)	-	-
	Louwerens et al (2014)	ND	0/20 (0%)	-	-
	Martimbianco et al (2020)	4/4 (100%)	1/4 (25%)	-	-
	Miller et al (2017)	ND	13/16 (81%)		
	Mohamadi et al (2017)	ND	4/14 (29%)	-	-
	Sussmilch-Leitch et al (2012)	4/23 (17%)	-	11/23 (48%)***	15/23 (65%)***

(Continues)

TABLE 2 (Continued)

Tool	SR	SR authors' "high overall RoB"	DC "high overall RoB" Cochrane Collaboration	DC "high overall RoB" PEDro	
	Tsikopoulos et al (2016)	ND	4/5 (80%)	-	-
	Toliopoulos et al (2014)	ND	7/15 (47%)	-	-
	Xiong et al (2019)	ND	0/4 (0%)	-	-
	Yan et al (2019)	ND	0/5 (0%)	-	-
	Zhang et al (2019)	ND	0/8 (0%)		
	Mean Proportion	34.6%	55%	-	-
RoB 2	Murphy et al (2019)	2/7 (29%)	NP**	-	-
	Van der Vlist et al (2020)	21/28 (75%)	NP**	-	-
	Mean Proportion	52%	-	-	-

Abbreviations: NA, not available; ND, not determined; SR, systematic review; RoB, risk of bias.

*Systematic review authors and author of present review (DC) used same criteria.

**Not performed as tool includes instructions on determination of overall risk of bias.

***Systematic review authors presented results of modified PEDro scale but assessed overall risk of bias based on Cochrane Collaboration tool.

4 | DISCUSSION

We have demonstrated several problems relating to the use of RoB assessment in SRs of tendinopathy management that need the attention of the research community. In our scoping review, we found that almost half of the included SRs did not use their RoB assessment in data synthesis. Additionally, only 6.5% of SRs assessed RoB on an outcome level and not a study level while only 30% of all SRs used their RoB assessment for evidence grading, which is the primary purpose of performing a RoB assessment. In light of the substantial subjectivity and lack of transparency and reproducibility that governs the conduct of SRs, we strongly recommend that future SR authors determine overall RoB for each study (on an outcome level) with the use of clear and reproducible pre-defined criteria.

Whether overall RoB should be determined or not for each RCT is a controversial question and this controversy is apparent in the tools themselves. Although the creators of the original Cochrane Collaboration tool[4] advised against rating overall RoB for each study but determining overall RoB on a domain level instead, this was neither explained further with clear, reproducible instructions nor was it applicable in practice for evidence grading. The revised Cochrane Collaboration tool (RoB 2)[5] published last year includes instructions on determining overall RoB for each study; however, the creators highlight that this needs to be done on an outcome level. Finally, the PEDro scale,[6,7] which its creators define as "a scale to measure the quality of reports of RCTs," does not define specific criteria or score cutoffs and is often incorrectly labeled as a "quality assessment" and not "RoB" tool. In addition to internal validity (RoB), measures of study quality include external validity (generalizability) and precision (freedom from random error), which the 10-item scale does not include. This is also acknowledged by the creators themselves.[7]

The comparison of the likelihood of each one of the three tools rating an RCT as "high overall risk" demonstrated clearly that the PEDro was overly generous as used by the SR authors, rating the majority of assessed RCTs (81.7%) as "low overall RoB"/"good overall quality." The possibility of that substantial proportion of tendinopathy RCTs actually being of "low overall RoB" is not even entertained; many of them are not double-blinded (due to their nature) and besides, the other two RoB assessment tools demonstrated greater proportions of "high overall RoB" RCTs. Finally, inter-tool reliability among the three tools was generally poor except for the comparison of the Cochrane Collaboration tool and the PEDro $\geq$ 8, which reinforces the need for PEDro to be used with stricter criteria.

When we assessed our own pre-defined criteria against those used by the SR authors, it was apparent that especially for the Cochrane Collaboration tool there were substantial discrepancies. One might argue that our strict criteria resulted in a very low threshold of rating an RCT as "high overall RoB"; however, the recently published RoB 2 is very close to our criteria in that respect as all it takes for a "high overall RoB" is high RoB in a single domain. These marked disparities reflect the significant effects that subjectivity, inconsistency, and lack of reproducibility can have on the results of the same SRs with regard to grading the quality of evidence. If we demonstrated inconsistencies this significant only by using different criteria for RoB assessment results as reported by the SR authors, one can imagine how much more substantial these disparities can be when the same RCTs are assessed by different people, with different tools, using different criteria for each tool. Finally, a naturally arising question is therefore "how much bias is enough to distort the true

TABLE 3A Our risk of bias assessment of the 29 RCTs included in the systematic review by van der Vlist (2020)[7] using the Cochrane Collaboration tool

	Internal Validity (Cochrane's Collaboration Tool for Assessing Risk of Bias)							
	Selection bias		Performance bias	Detection bias	Attrition bias	Reporting bias		
First Author (y)	Random sequence generation	Allocation concealment	Blinding of patients and staff	Blinding of outcome measures	Completeness of outcome data	Selective reporting	Other	Overall RoB
Balius et al (2016)	Low	?	High	High	Low	Low	Low	High
Bell et al (2013)	Low	Low	Low	Low	Low	Low	Low	Low
Beyer et al (2015)	Low	?	High	High	?	Low	?	High
Boesen et al (2017)	Low	Low	Low	Low	Low	Low	Low	Low
De Jonge et al (2010)	?	Low	High	High	High	Low	Low	High
De Jonge et al (2011)	Low	Low	Low	Low	Low	Low	Low	Low
Ebbesen et al (2017)	?	Low	Low	Low	Low	Low	High	Low
Heinemeier et al (2017)	Low	Low	Low	Low	?	Low	?	Low
Herrington & McCulloch (2007)	High	?	High	High	Low	Low	Low	High
Hutchison et al (2013)	Low	Low	Low	Low	Low	Low	Low	Low
Krogh et al (2016)	Low	Low	Low	High	Low	Low	Low	High
Lynen et al (2017)	Low	Low	High	High	Low	Low	Low	High
Morrison et al (2017)	Low	Low	Low	High	Low	Low	Low	High
Munteanu et al (2015)	Low	Low	Low	Low	Low	Low	Low	Low
Njawaya et al (2018)	Low	Low	High	High	Low	High	High	High
Pearson et al (2012)	Low	?	High	High	High	High	High	High
Rompe et al (2008)	Low	Low	High	High	Low	Low	Low	High
Rompe et al (2009)	Low	Low	High	High	Low	Low	High	High
Rompe et al (2009)	Low	Low	High	High	Low	Low	Low	High
Roos et al (2004)	Low	?	High	High	High	Low	High	High
Silbernagel et al (2001)	?	?	High	High	High	High	High	High
Silbernagel et al (2007)	Low	Low	High	High	?	High	Low	High
Stevens & Tan (2014)	?	Low	High	High	High	Low	Low	High
Tumilty et al (2016)	Low	Low	Low	Low	High	Low	Low	Low
Tumilty et al (2012)	Low	Low	Low	Low	Low	Low	Low	Low
Usuelli et al (2018)	?	Low	High	High	Low	High	Low	High
Yelland et al (2009)	Low	?	High	High	Low	Low	High	High
Zhang et al (2013)	Low	Low	High	High	Low	Low	Low	High

result of an RCT?"; unfortunately, this and other similarly subjective judgments are needed for the conduct and reporting of all SRs.

The ideal RoB assessment tool does not exist. Subjectivity can never be removed completely from RoB assessment; however, this needs to be kept to a minimum and be complemented by transparency and reproducibility. These are exactly the aims of the revised Cochrane Collaboration tool, the creators of which state that they expect the new tool to be more likely to rate studies as "low overall RoB."[5] This was clearly not the case with the example SR used in the present review by van der Vlist et al[8] who rated none of the 29 RCTs as "low risk." Reasons for that might be either the actual presence of bias in all the included RCTs, strict thresholds used by the SR authors or poor performance of the tool itself. The same tool applied in the other SR[46] included in this review identified a much higher proportion of "low overall RoB" RCTs (4/7). Despite attempts of the creators to make the tool more user friendly and reproducible,[4] there is still significant subjectivity in some of its signaling questions (eg, "could assessment have been influenced by knowledge of intervention?" or "likely that missingness depended on true value"). However, importantly the

TABLE 3B Our risk of bias assessment of the 29 RCTs included in the systematic review by van der Vlist (2020)[7] using the PEDro Tool

Study	1	2	3	4	5	6	7	8	9	10	Total Score	Overall ≥ 6	Overall ≥ 8
Balius et al (2016)	Yes	No	Yes	No	No	No	Yes	Yes	Yes	Yes	6	Low	High
Bell et al (2013)	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	9	Low	Low
Beyer et al (2015)	Yes	No	Yes	No	No	No	No	No	Yes	Yes	4	High	High
Boesen et al (2017)	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	9	Low	Low
De Jonge et al (2010)	Yes	Yes	Yes	No	No	No	No	Yes	Yes	No	5	**High**	High
De Jonge et al (2011)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	10	Low	Low
Ebbesen et al (2017)	No	Yes	Yes	Yes	No	Yes	Yes	Yes	No	Yes	7	Low	High
Heinemeier et al (2017)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	10	Low	Low
Herrington & McCulloch (2007)	No	No	Yes	No	No	No	Yes	Yes	Yes	No	4	High	High
Hutchison et al (2013)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	10	Low	Low
Krogh et al (2016)	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes	8	Low	Low
Lynen et al (2017)	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes	Yes	7	Low	High
Morrison et al (2017)	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes	8	Low	Low
Munteanu et al (2015)	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	9	Low	Low
Njawaya et al (2018)	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes	No	6	Low	High
Pearson et al (2012)	Yes	No	Yes	No	No	No	No	Yes	No	No	3	High	High
Rompe et al (2008)	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes	Yes	7	**Low**	High
Rompe et al (2009)	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes	Yes	7	Low	High
Rompe et al (2007)	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes	Yes	7	Low	High
Roos et al (2004)	Yes	No	No	No	No	No	No	Yes	Yes	Yes	4	High	High
Silbernagel et al (2001)	Yes	No	No	No	No	No	No	Yes	No	Yes	3	High	High
Silbernagel et al (2007)	Yes	Yes	Yes	No	No	No	No	Yes	No	Yes	5	High	High
Stevens & Tan (2014)	No	Yes	Yes	No	No	No	No	Yes	Yes	Yes	5	High	High
Tumilty et al (2016)	Yes	Yes	Yes	Yes	No	Yes	No	Yes	Yes	Yes	8	Low	Low
Tumilty et al (2012)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	10	Low	Low
Usuelli et al (2018)	No	Yes	Yes	No	No	No	Yes	Yes	Yes	No	5	High	High
Yelland et al (2009)	Yes	Yes	No	No	No	No	Yes	Yes	Yes	Yes	6	Low	High
Zhang et al (2013)	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes	Yes	7	Low	**High**

tool includes clear instructions on determining both RoB for each individual domain and overall RoB for each study and this is why we advocate its use by all future SR authors.

4.1 | Recommendations

In order to minimize inconsistency in RoB assessment and its use in data synthesis, we suggest the consistent use of RoB assessment across all journals publishing SRs. This will be achieved through the use of a single RoB assessment tool that can be incorporated in the "Instructions for authors" section of each journal's website or even in the PRISMA statement[57] and other SR guidance documents. Additionally, for subjectivity and lack of transparency to be kept to a minimum, RCT authors could include a RoB assessment of their own study (with justifications) that will remove the need for authors' judgments at an SR level. Similarly, this could be achieved by the consistent use of the same tool across publishing journals and its introduction in RCT guidance documents (eg, CONSORT).[58] Finally, journals and reviewers should apply more stringent criteria for accepting low-quality RCTs and SRs with inadequate transparency and reproducibility.

5 | CONCLUSION

In the present review, we demonstrate several issues regarding the use of RoB assessment in tendinopathy SRs both relating to the tools themselves and their use by authors. Most importantly, there appears to be a lack of understanding on

TABLE 3C van der Vlist's[7] RoB assessment of the 29 included RCTs using the RoB 2 and comparison with our assessment

Study	Randomization	Deviations from protocol	Missing data	Measurement of outcome	Selection of result	Overall RoB	Cochrane (DC)	PEDro ≥ 6 (DC)	PEDro ≥ 8 (DC)
Balius et al (2016)	High	Some concerns	Low	High	High	High	High	Low	High
Bell et al (2013)	Low	Low	Low	Low	Some concerns	Some concerns	Low	Low	Low
Beyer et al (2015)	Some concerns	Some concerns	High	High	High	High	High	High	High
Boesen et al (2017)	High	Some concerns	Low	Low	Some concerns	High	Low	Low	Low
De Jonge et al (2010)	Some concerns	Some concerns	Low	High	Some concerns	High	High	High	High
De Jonge et al (2011)	Some concerns	Low	Low	Low	Some concerns	Some concerns	Low	Low	Low
Ebbesen et al (2017)	High	High	Some concerns	Low	High	High	Low	Low	High
Heinemeier et al (2017)	Some concerns	Low	Low	Low	Some concerns	Some concerns	Low	Low	Low
Herrington & McCulloch (2007)	Some concerns	Low	Low	High	Some concerns	High	High	High	High
Hutchison et al (2013)	High	High	High	Low	High	High	Low	Low	Low
Krogh et al (2016)	Some concerns	High	High	Low	Some concerns	High	High	Low	Low
Lynen et al (2017)	Low	Low	Some concerns	High	High	High	High	Low	High
Morrison et al (2017)	High	Low	Low	Low	Some concerns	High	High	Low	Low
Munteanu et al (2015)	Low	High	Low	Low	Low	High	Low	Low	Low
Njawaya et al (2018)	Some concerns	Low	Some concerns	High	High	High	High	Low	High
Pearson et al (2012)	Some concerns	Some concerns	High	High	Some concerns	High	High	High	High
Rompe et al (2008)	Low	High	High	High	Some concerns	High	High	Low	High
Rompe et al (2009)	Low	Some concerns	High	High	Some concerns	High	High	Low	High
Rompe et al (2007)	Low	Low	Some concerns	High	Some concerns	High	High	Low	High
Roos et al (2004)	Some concerns	High	Some concerns	Low	Some concerns	High	High	High	High
Silbernagel et al (2001)	Some concerns	High	High	Some concerns	Some concerns	High	High	High	High
Silbernagel et al (2007)	Some concerns	Some concerns	Some concerns	Low	Some concerns	Some concerns	High	High	High
Stevens & Tan (2014)	Some concerns	Some concerns	Low	Some concerns	Some concerns	Some concerns	High	High	High
Tumilty et al (2016)	Some concerns	Some concerns	Some concerns	High	Low	High	Low	Low	Low
Tumilty et al (2012)	Low	Low	Some concerns	Low	Some concerns	Some concerns	Low	Low	Low
Usuelli et al (2018)	Some concerns	Low	Low	High	Some concerns	High	High	High	High
Yelland et al (2009)	Some concerns	Low	Some concerns	Some concerns	Some concerns	Some concerns	High	Low	High
Zhang et al (2013)	Some concerns	Some concerns	Some concerns	High	Some concerns	High	High	Low	High
Total Overall RoB	-	-	-	-	-	0 low, 7 some concerns, 21 high	9 Low, 19 High, 0 unclear	19 Low, 9 High	18 High, 10 Low

Note: DC, as determined by first author of present review.

the appropriate use of RoB assessment and its incorporation in data syntheses. We recommend the consistent use of a single RoB assessment tool across all publishing journals and guidance documents and the application of more stringent criteria when both RCTs and SRs are assessed for publication.

CONFLICT OF INTERESTS

The authors declare no competing financial interests.

AUTHOR CONTRIBUTIONS

DC and NLM conceived and designed the study, performed analysis, and wrote the manuscript. All authors analyzed the data.

DATA AVAILABILITY STATEMENT

DC has access to all the data, and data are available upon request.

ORCID

Neal L. Millar https://orcid.org/0000-0001-9251-9907

REFERENCES

1. National Institute for Health Research (NIHR). Evidence synthesis and systematic reviews. https://www.journalslibrary.nihr.ac.uk/information-for-authors/evidence-synthesis-and-systematic-reviews/. Accessed June 15, 2020
2. Horton R. Offline: The gravy train of systematic reviews. *Lancet*. 2019;394(10211):1790.
3. Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*. 2008;336:924.
4. Higgins JP, Altman DG, Gøtzsche PC, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ*. 2011;343:d5928.
5. Sterne JAC, Savović J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ*. 2019;366:l4898.
6. Verhagen AP, de Vet HC, de Bie RA, et al. The Delphi list: a criteria list for quality assessment of randomized clinical trials for conducting systematic reviews developed by Delphi consensus. *J Clin Epidemiol*. 1998;51(12):1235-1241.
7. PEDro Scale. Physiotherapy evidence database. http://www.pedro.org.au/English/downloads/pedro-scale/. Accessed June 9, 2020
8. van der Vlist AC, Winters M, Weir A, et al. Which treatment is most effective for patients with Achilles tendinopathy? A living systematic review with network meta-analysis of 29 randomised controlled trials Published Online First: 10 June 2020. doi: 10.1136/bjsports-2019-101872. *Br J Sports Med*. 2020;bjsports-2019-101872.
9. van Tulder M, Furlan A, Bombardier C, Bouter L. Editorial Board of the Cochrane Collaboration Back Review Group. Updated method guidelines for systematic reviews in the cochrane collaboration back review group. *Spine (Phila Pa 1976)*. 2003;28(12):1290-1299.
10. Chalmers TC, Celano P, Sacks HS, et al. Bias in treatment assignment in controlled clinical trials. *N Engl J Med*. 1983;309:1358-1361.
11. Kunz R, Oxman AD. The unpredictability paradox: review of empirical comparisons of randomised and non-randomised clinical trials. *BMJ*. 1998;317:1185-1190.
12. Arirachakaran A, Sukthuayat A, Sisayanarane T, Laoratanavoraphong S, Kanchanatawan W, Kongtharvonskul J. Platelet-rich plasma versus autologous blood versus steroid injection in lateral epicondylitis: systematic review and network meta-analysis. *J Orthop Traumatol*. 2016;17(2):101-112.
13. Arirachakaran A, Boonard M, Yamaphai S, Prommahachai A, Kesprayura S, Kongtharvonskul J. Extracorporeal shock wave therapy, ultrasound-guided percutaneous lavage, corticosteroid injection and combined treatment for the treatment of rotator cuff calcific tendinopathy: a network meta-analysis of RCTs. *Eur J Orthop Surg Traumatol*. 2017;27(3):381-390.
14. Bannuru RR, Flavin NE, Vaysbrot E, Harvey W, McAlindon T. High-energy extracorporeal shock-wave therapy for treating chronic calcific tendinitis of the shoulder: a systematic review. *Ann Intern Med*. 2014;160(8):542-549.
15. Bjordal JM, Lopes-Martins RA, Joensen J, et al. A systematic review with procedural assessments and meta-analysis of low level laser therapy in lateral elbow tendinopathy (tennis elbow). *BMC Musculoskelet Disord*. 2008;9:75.
16. Boudreault J, Desmeules F, Roy JS, Dionne C, Frémont P, Macdermid JC. The efficacy of oral non-steroidal anti-inflammatory drugs for rotator cuff tendinopathy: a systematic review and meta-analysis. *J Rehabil Med*. 2014;46(4):294-306.
17. Catapano M, Catapano J, Borschel G, Alavinia SM, Robinson LR, Mittal N. Effectiveness of platelet-rich plasma injections for nonsurgical management of carpal tunnel syndrome: a systematic review and meta-analysis of randomized controlled trials. *Arch Phys Med Rehabil*. 2020;101(5):897-906.
18. Challoumas D, Clifford C, Kirwan P, Millar NL. How does surgery compare to sham surgery or physiotherapy as a treatment for tendinopathy? A systematic review of randomised trials. *BMJ Open Sport Exerc Med*. 2019;5(1):e000528.
19. Challoumas D, Kirwan PD, Borysov D, Clifford C, McLean M, Millar NL. Topical glyceryl trinitrate for the treatment of tendinopathies: a systematic review. *Br J Sports Med*. 2019;53(4):251-262.
20. Chen PC, Wu KT, Chou WY, et al. Comparative effectiveness of different nonsurgical treatments for patellar tendinopathy: a systematic review and network meta-analysis. *Arthroscopy*. 2019;35(11):3117-3131.e2.
21. Coombes BK, Bisset L, Vicenzino B. Efficacy and safety of corticosteroid injections and other injections for management of tendinopathy: a systematic review of randomised controlled trials. *Lancet*. 2010;376(9754):1751-1767.
22. Dan M, Phillips A, Johnston RV, Harris IA. Surgery for patellar tendinopathy (Jumper's knee). *Cochrane Database Syst Rev*. 2019, Issue 9. Art. No.: CD013034.
23. de Vos RJ, Windt J, Weir A. Strong evidence against platelet-rich plasma injections for chronic lateral epicondylar tendinopathy: a systematic review. *Br J Sports Med*. 2014;48(12):952-956.
24. Desjardins-Charbonneau A, Roy JS, Dionne CE, Desmeules F. The efficacy of taping for rotator cuff tendinopathy: a systematic review and meta-analysis. *Int J Sports Phys Ther*. 2015;10(4):420-433.
25. Desjardins-Charbonneau A, Roy JS, Dionne CE, Frémont P, MacDermid JC, Desmeules F. The efficacy of manual therapy for rotator cuff tendinopathy: a systematic review and meta-analysis. *J Orthop Sports Phys Ther*. 2015;45(5):330-350.
26. Desmeules F, Boudreault J, Dionne CE, et al. Efficacy of exercise therapy in workers with rotator cuff tendinopathy: a systematic review. *J Occup Health*. 2016;58(5):389-403.

27. Desmeules F, Boudreault J, Roy JS, Dionne CE, Frémont P, MacDermid JC. Efficacy of transcutaneous electrical nerve stimulation for rotator cuff tendinopathy: a systematic review. *Physiotherapy*. 2016;102(1):41-49.
28. Desmeules F, Boudreault J, Roy JS, Dionne C, Frémont P, MacDermid JC. The efficacy of therapeutic ultrasound for rotator cuff tendinopathy: A systematic review and meta-analysis. *Phys Ther Sport*. 2015;16(3):276-284.
29. Dong W, Goost H, Lin XB, et al. Injection therapies for lateral epicondylalgia: a systematic review and Bayesian network meta-analysis. *Br J Sports Med*. 2016;50(15):900-908.
30. Dong W, Goost H, Lin XB, et al. Treatments for shoulder impingement syndrome: a PRISMA systematic review and network meta-analysis [published correction appears in Medicine (Baltimore). 2016 Jun 10;95(23):e96d5]. *Medicine (Baltimore)*. 2015;94(10):e510.
31. Fitzpatrick J, Bulsara M, Zheng MH. The effectiveness of platelet-rich plasma in the treatment of tendinopathy: a meta-analysis of randomized controlled clinical trials. *Am J Sports Med*. 2017;45(1):226-233.
32. Haslerud S, Magnussen LH, Joensen J, Lopes-Martins RA, Bjordal JM. The efficacy of low-level laser therapy for shoulder tendinopathy: a systematic review and meta-analysis of randomized controlled trials. *Physiother Res Int*. 2015;20(2):108-125.
33. Ioppolo F, Tattoli M, Di Sante L, et al. Clinical improvement and resorption of calcifications in calcific tendinitis of the shoulder after shock wave therapy at 6 months' follow-up: a systematic review and meta-analysis. *Arch Phys Med Rehabil*. 2013;94(9):1699-1706.
34. Lafrance S, Doiron-Cadrin P, Saulnier M, et al. Is ultrasound-guided lavage an effective intervention for rotator cuff calcific tendinopathy? A systematic review with a meta-analysis of randomised controlled trials. *BMJ Open Sport Exerc Med*. 2019;5(1):e000506.
35. Lee SY, Cheng B, Grimmer-Somers K. The midterm effectiveness of extracorporeal shockwave therapy in the management of chronic calcific shoulder tendinitis. *J Shoulder Elbow Surg*. 2011;20(5):845-854.
36. Li A, Wang H, Yu Z, et al. Platelet-rich plasma vs corticosteroids for elbow epicondylitis: a systematic review and meta-analysis. *Medicine (Baltimore)*. 2019;98(51):e18358.
37. Liao CD, Tsauo JY, Chen HC, Liou TH. Efficacy of extracorporeal shock wave therapy for lower-limb tendinopathy: a meta-analysis of randomized controlled trials. *Am J Phys Med Rehabil*. 2018;97(9):605-619.
38. Lin MT, Wei KC, Wu CH. Effectiveness of platelet-rich plasma injection in rotator cuff tendinopathy: a systematic review and meta-analysis of randomized controlled trials. *Diagnostics (Basel)*. 2020;10(4):189.
39. Lin MT, Chiang CF, Wu CH, Huang YT, Tu YK, Wang TG. Comparative effectiveness of injection therapies in rotator cuff tendinopathy: a systematic review, pairwise and network meta-analysis of randomized controlled trials. *Arch Phys Med Rehabil*. 2019;100(2):336-349.e15.
40. Lin YC, Wu WT, Hsu YC, Han DS, Chang KV. Comparative effectiveness of botulinum toxin versus non-surgical treatments for treating lateral epicondylitis: a systematic review and meta-analysis. *Clin Rehabil*. 2018;32(2):131-145.
41. Louwerens JK, Sierevelt IN, van Noort A, van den Bekerom MP. Evidence for minimally invasive therapies in the management of chronic calcific tendinopathy of the rotator cuff: a systematic review and meta-analysis. *J Shoulder Elbow Surg*. 2014;23(8):1240-1249.
42. Martimbianco ALC, Ferreira RES, Latorraca COC, Bussadori SK, Pacheco RL, Riera R. Photobiomodulation with low-level laser therapy for treating Achilles tendinopathy: a systematic review and meta-analysis. *Clin Rehabil*. 2020;34(6):713-722.
43. Mendonça LM, Leite HR, Zwerver J, Henschke N, Branco G, Oliveira VC. How strong is the evidence that conservative treatment reduces pain and improves function in individuals with patellar tendinopathy? A systematic review of randomised controlled trials including GRADE recommendations. *Br J Sports Med*. 2020;54(2):87-93.
44. Miller LE, Parrish WR, Roides B, Bhattacharyya S. Efficacy of platelet-rich plasma injections for symptomatic tendinopathy: systematic review and meta-analysis of randomised injection-controlled trials. *BMJ Open Sport Exerc Med*. 2017;3(1):e000237.
45. Mohamadi A, Chan JJ, Claessen FM, Ring D, Chen NC. Corticosteroid injections give small and transient pain relief in rotator cuff tendinosis: a meta-analysis. *Clin Orthop Relat Res*. 2017;475(1):232-243.
46. Murphy MC, Travers MJ, Chivers P, et al. Efficacy of heavy eccentric calf training for treating mid-portion Achilles tendinopathy: a systematic review and meta-analysis. *Br J Sports Med*. 2019;53(17):1070-1077.
47. Ortega-Castillo M, Medina-Porqueres I. Effectiveness of the eccentric exercise therapy in physically active adults with symptomatic shoulder impingement or lateral epicondylar tendinopathy: a systematic review. *J Sci Med Sport*. 2016;19(6):438-453.
48. Sussmilch-Leitch SP, Collins NJ, Bialocerkowski AE, Warden SJ, Crossley KM. Physical therapies for Achilles tendinopathy: systematic review and meta-analysis. *J Foot Ankle Res*. 2012;5(1):15.
49. Tsikopoulos K, Tsikopoulos A, Natsis K. Autologous whole blood or corticosteroid injections for the treatment of epicondylopathy and plantar fasciopathy? A systematic review and meta-analysis of randomized controlled trials. *Phys Ther Sport*. 2016;22:114-122.
50. Toliopoulos P, Desmeules F, Boudreault J, et al. Efficacy of surgery for rotator cuff tendinopathy: a systematic review. *Clin Rheumatol*. 2014;33(10):1373-1383.
51. Wasielewski NJ, Kotsko KM. Does eccentric exercise reduce pain and improve strength in physically active adults with symptomatic lower extremity tendinosis? A systematic review. *J Athl Train*. 2007;42(3):409-421.
52. Woodley BL, Newsham-West RJ, Baxter GD. Chronic tendinopathy: effectiveness of eccentric exercise. *Br J Sports Med*. 2007;41(4):188-199.
53. Wu YC, Tsai WC, Tu YK, Yu TY. Comparative effectiveness of nonoperative treatments for chronic calcific tendinitis of the shoulder: a systematic review and network meta-analysis of randomized controlled trials. *Arch Phys Med Rehabil*. 2017;98(8):1678-1692.e6.
54. Xiong Y, Xue H, Zhou W, et al. Shock-wave therapy versus corticosteroid injection on lateral epicondylitis: a meta-analysis of randomized controlled trials. *Phys Sportsmed*. 2019;47(3):284-289.
55. Yan C, Xiong Y, Chen L, et al. A comparative study of the efficacy of ultrasonics and extracorporeal shock wave in the treatment of tennis elbow: a meta-analysis of randomized controlled trials. *J Orthop Surg Res*. 2019;14(1):248. Published 2019 Aug 6.
56. Zhang T, Duan Y, Chen J, Chen X. Efficacy of ultrasound-guided percutaneous lavage for rotator cuff calcific tendinopathy: a

systematic review and meta-analysis. *Medicine (Baltimore)*. 2019; 98(21):e15552.
57. Moher D, Liberati A, Tetzlaff J, Altman DG, for the PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med*. 2009;6(7): e1000097.
58. Schulz KF, Altman DG, Moher D, for the CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *Ann Int Med*. 2010;2010:152.

How to cite this article: Challoumas D, Millar NL. Risk of bias in systematic reviews of tendinopathy management: Are we comparing apples with oranges? *Transl Sports Med*. 2021;4:21–37. https://doi.org/10.1002/tsm2.196