

# Do we need to improve the reporting of evidence in tendinopathy management? A critical appraisal of systematic reviews with recommendations on strength of evidence assessment

Dimitris Challoumas, Neal L Millar 

**To cite:** Challoumas D, Millar NL. Do we need to improve the reporting of evidence in tendinopathy management? A critical appraisal of systematic reviews with recommendations on strength of evidence assessment. *BMJ Open Sport & Exercise Medicine* 2021;**7**:e000920. doi:10.1136/bmjsem-2020-000920

► Additional material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/bmjsem-2020-000920>).

Accepted 2 January 2021



© Author(s) (or their employer(s)) 2021. Re-use permitted under CC BY. Published by BMJ.

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

## Correspondence to

Mr Neal L Millar;  
neal.millar@glasgow.ac.uk

## ABSTRACT

**Objective** To critically appraise the quality of published systematic reviews (SRs) of randomised controlled trials (RCTs) in tendinopathy with regard to handling and reporting of results with special emphasis on strength of evidence assessment.

**Data sources** Medline from inception to June 2020.

**Study eligibility** All SRs of RCTs assessing the effectiveness of any intervention(s) on any location of tendinopathy.

**Data extraction and synthesis** Included SRs were appraised with the use of a 12-item tool devised by the authors arising from the Preferred Reporting Items in Systematic Reviews and Meta-Analyses statement and other relevant guidance. Subgroup analyses were performed based on impact factor (IF) of publishing journals and date of publication.

**Results** A total of 57 SRs were included published in 38 journals between 2006 and 2020. The most commonly used risk-of-bias (RoB) assessment tool and strength of evidence assessment tool were the Cochrane Collaboration RoB tool and the Cochrane Collaboration Back Review Group tool, respectively. The mean score on the appraisal tool was 46.5% (range 0%–100%). SRs published in higher IF journals (>4.7) were associated with a higher mean score than those in lower IF journals (mean difference 26.4%±8.8%,  $p=0.004$ ). The mean score of the 10 most recently published SRs was similar to that of the first 10 published SRs (mean difference 8.3%±13.7%,  $p=0.54$ ). Only 23 SRs (40%) used the results of their RoB assessment in data synthesis and more than half ( $n=30$ ; 50%) did not assess the strength of evidence of their results. Only 12 SRs (21%) assessed their strength of evidence appropriately.

**Conclusions** In light of the poor presentation of evidence identified by our review, we provide recommendations to increase transparency and reproducibility in future SRs.

## INTRODUCTION

The ever-expanding arsenal of treatment regimes for tendinopathy can be overwhelming to the treating healthcare professional. New treatments continuously

## Summary box

### What is already known?

- The increasing number of available treatment modalities for tendinopathy can be overwhelming to the treating healthcare professionals.
- Systematic reviews (SRs) of randomised controlled trials provide the highest level of evidence and guide application of research findings to clinical practice.
- There are several strength of evidence assessment tools used in tendinopathy SRs.
- The use of strength of evidence assessment results in data synthesis is associated with confusion and inconsistency.

### What are the new findings?

- More than half of the SRs did not assess the strength of evidence in their results.
- Only 21% of SRs assessed their strength of evidence appropriately.
- Authors who publishing SRs on the management of tendinopathy should become familiar with strength of evidence assessment tools and apply them appropriately in their data synthesis and presentation.
- We provide recommendations to increase transparency and reproducibility in future SRs.

emerge and so do research studies that aim to assess their effectiveness. The most powerful tool that evidence-based medicine has to offer remains systematic reviews (SRs) of randomised controlled trials (RCTs), which constitute the highest level of evidence and are therefore often used to inform guidelines and guide clinical practice.<sup>1</sup>

Unlike a narrative review, which is purely a summary of a proportion of (or all) the available studies on a given topic without assessing the quality of the included evidence, an SR is expected to include all the relevant evidence which is comprehensively and thoroughly appraised and presented with qualitative or quantitative syntheses and accompanying

strengths of evidence.<sup>2</sup> For reliable, clinically relevant conclusions to be reached, the potential limitations of the included studies need to be considered.<sup>3</sup> What determines the ‘quality’ of a study remains poorly defined but should relate to the extent to which its design, conduct, analysis and presentation are appropriate to answer its research question.<sup>3</sup> Generally, study quality assessment includes a combined assessment of its internal validity (freedom from bias), external validity (generalisability/applicability) and precision (freedom from random error)<sup>3</sup> (online supplemental figure 1). The Cochrane group, which produce some of the highest quality SRs, recommend emphasising the internal validity when assessing the quality of a study, reflected by the risk-of-bias assessment (RoB) tool they have produced.<sup>3</sup> However, external validity and precision should not be overlooked.

Assessing the strength of evidence (also known as ‘certainty’, ‘quality’, ‘level’ or ‘grade’ of evidence) is closely interlinked with RoB assessment and is a process which is commonly overlooked by many SRs. Together with a meticulous methodology and comprehensive RoB assessment, this is what distinguishes a SR from other types of review articles. Assigning a strength of evidence to a finding is as important as the finding itself. Strength of evidence should accompany every assessed outcome measure within an SR for every assessed follow-up time period.<sup>4,5</sup> A result with a low strength of evidence differs substantially from that with a high level of evidence in terms of its applicability to clinical practice; further research is likely to change the former but not the latter.<sup>6</sup> Equally, the strength of evidence can be high in an outcome measure and low in another, and high for a certain follow-up but period and low for another.

An SR should be transparent and reproducible, and subjectivity should be kept to a minimum.<sup>7</sup> Firm guidance on conducting SRs does not exist and several parameters are left to the judgement of the authors. In an attempt to optimise the quality of SRs, the Preferred Reporting Items in Systematic Reviews and Meta-Analyses (PRISMA) group have published evidence-based guidance, including a 27-item checklist.<sup>8</sup> Similarly, the A MeaSurement Tool to Assess systematic Reviews (AMSTAR) group have published a critical appraisal tool for SRs aiming to help the reader/appraiser make conclusions about its quality.<sup>9</sup> The aim of the former in particular is to help authors improve the reporting of SRs and meta-analyses of randomised trials and it is considered the gold standard for the authorship of SRs. However, summarising the strength of the presented evidence, which is mentioned in one of its items, does not receive sufficient attention in most SRs. 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.<sup>10</sup>

Our aim was to critically appraise the quality of published SRs of RCTs in tendinopathy, with emphasis on handling and reporting of results and especially assessment of the strength of evidence, which is often

overlooked. We provide recommendations for authors of future SRs aiming to optimise data synthesis, reporting and summarising the strength of evidence. The present review is also intended for readers of SRs, who need to critically appraise their quality before reaching conclusions about their findings, and finally it can be used by reviewers and editors of journals when assessing submitted SRs for publication.

## METHODS

### Eligibility

Eligible articles assessed the effectiveness of any intervention(s) on tendinopathy (any type), identified themselves as an SR and/or meta-analysis and only included RCTs. Reviews including a mixture of randomised studies and other types of studies (non-randomised clinical trials, cohort studies, case series, etc) and those assessing the effectiveness of interventions on tendinopathy and other conditions were excluded. Only articles published in English were included in patients over 16 years of age. No criteria were used regarding the following parameters: publication date, journal type, type of tendinopathy and intervention, outcome measures and length of follow-up.

### 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 flow chart) illustrates the article screening process.

The initial search returned a total of 207 articles. After exclusion of non-eligible articles according to our predefined criteria and inclusion of articles identified from reference screening, 57 reviews were included in our appraisal.

### Data extraction: handling

The included articles were read by the first author and their key characteristics were tabulated in Microsoft Word. Each article was then reread and appraised with regard to the handling and reporting of results based on our prespecified appraisal tool.

### Appraisal tool: rationale and explanation of items

We developed a 12-item appraisal checklist based on guidance from the PRISMA checklist predominantly, as well as other key documents, namely the Grading of Recommendations Assessment, Development and Evaluation (GRADE) tool, the AMSTAR checklist and the Cochrane Collaboration Back Review Group.<sup>4 5 8 9</sup> We highlight that our tool was created merely for the purposes of the present review, it does not intend to appraise SRs as a whole and it cannot replace other tools such as the PRISMA checklist and the AMSTAR tool; it can, however,

be used in conjunction with them. It intends to assess data synthesis and presentation without considering other important methodological aspects (eg, search strategy, screening, inclusion criteria). Each item receives either a ‘0’, ‘1’ or ‘2’ except for four items (2, 4, 11, 12) which only receive ‘0’ or ‘1’. The maximum score is 20 for SRs that include both qualitative and quantitative analyses and 18 for those including one or the other. The tool is intended to be used as a checklist rather than a scoring system, however, scoring was used in the present review for the purposes of subgroup comparisons. The 12 items are divided into the following three categories and a description of how the tool was used is provided below:

**ROB assessment**

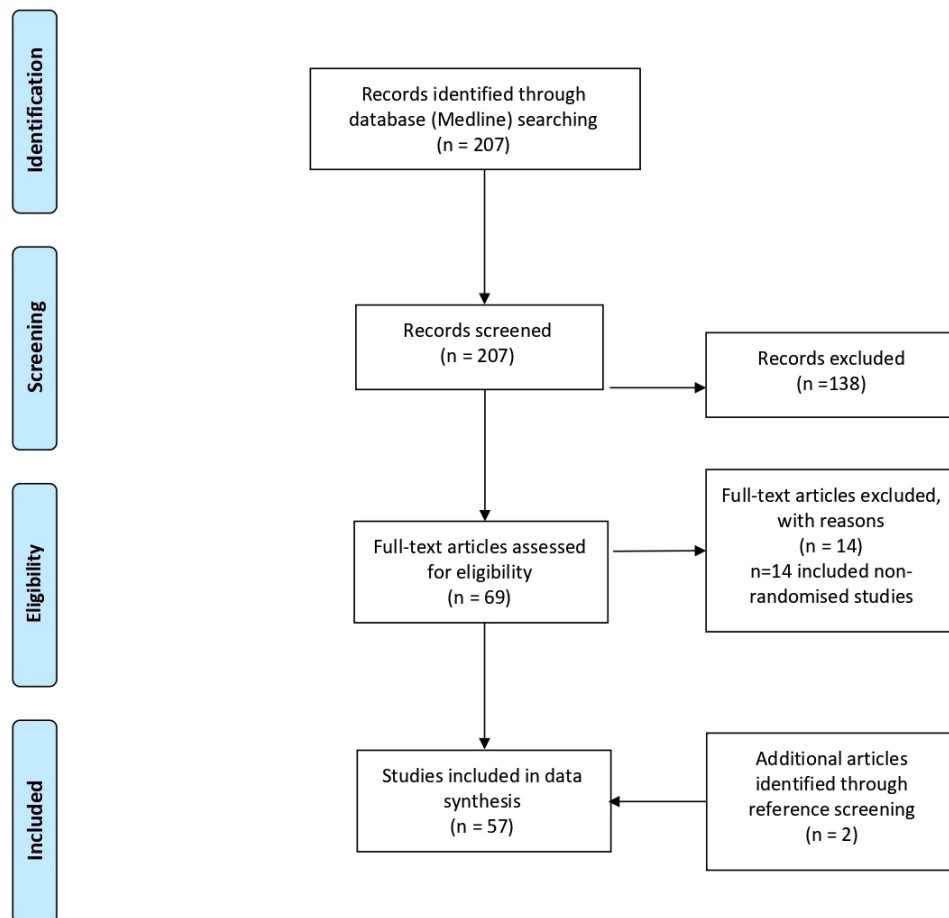
Item 1: The RoB assessment of each included RCTs, with or without an ‘overall risk’ for each study, needs to be used somehow for summarising the strength of evidence, which is its primary purpose, and this needs to be explained in the methods. This can be through subgroup analyses, for example, only synthesising results of studies with low overall RoB (which is indirectly associated with strength of evidence assessment).

Item 2; RoB should be assessed for each outcome measure separately to determine the strength of

evidence of the results for that specific outcome measure. Assessing RoB on an outcome measure level rather than a study level becomes particularly important when the study includes patient-reported measures and measures assessed directly by another assessor. With the former, if the participants are not blinded then automatically the assessment of outcome measures cannot be blinded either as the assessor is the participant his/herself. For non-patient-reported measures, regardless of blinding of participants, the blinding of assessment depends solely on whether the assessor (researcher) is blinded or not. If RoB is assessed on a study and not outcome level this should be stated in the methods with a justification (ie, all outcome measures were patient reported or participants of all studies blinded, etc). For SRs including one outcome measure only, this item was not scored.

**Pooling of results**

Item 3: The principal summary measures for each included outcome measure need to be stated in the methods. For quantitative analyses, where outcome measures are continuous, a justification needs to be provided for the use of (raw) mean differences (MD) and not standardised MD (SMD) and vice versa (ie, identical or different outcome measure tools used across studies,



**Figure 1** PRISMA flow diagram of included studies. PRISMA, Preferred Reporting Items in Systematic Reviews and Meta-Analyses.

respectively). Similarly, for dichotomous outcomes, the authors should state in the methods whether OR or relative risks (RR) were used.

Item 4: An SR without a meta-analysis should ideally have some quantitative summary measures (ie, MD, SMD, OR or RR with accompanying CI) to demonstrate the treatment effect of the assessed intervention over the comparator. Where results are pooled only based on direction of effect (ie, increased, decreased or unchanged), this needs to be stated in the methods with a justification. If pooling is not possible and results are only described narratively, the reason should be stated (usually substantial clinical heterogeneity); in such cases, we question whether the article should be identified as an SR.

Item 5: When meta-analyses are performed, the model used needs to be stated (ie, fixed effects or random effects) with a justification and the statistical heterogeneity assessment (usually  $\chi^2$  and/or  $I^2$  test) and how it was used in data syntheses. Sensitivity and subgroup analyses should also be pre-defined in the methods.

Items 6 and 7: Results should only be pooled for similar follow-up time periods which should be prespecified in the methods. Traditionally these are short term, mid term and long term. The range of these should be defined by the authors in the methods based on the population, outcome measures and interventions. Equally, assessed outcome measures should be predefined (usually as part of PICOS).

Item 8: A statement needs to be included in the section explaining how the authors dealt with missing data, which are usually methodological details of the included studies needed for the RoB assessment and descriptive statistics for quantitative analyses. For both, ideally attempts should be made to contact the authors of the RCTs for retrieval of the missing data and this should be stated. If significant statistical data of included studies cannot be retrieved (ie, sample sizes and means) these studies should be excluded from quantitative analyses; if variability statistics (SD) are missing, the authors have the option to impute these using data from other RCTs and the imputation method used should be described.

#### Strength of evidence assessment

Items 9–12: Assigning a strength of evidence to a result is an essential part of an SR and it should be determined separately for each predefined outcome measure at each predefined follow-up time period. By and large this is determined based on the overall quality of the included studies and the extent of consistency of their findings with considerations of other limitations across studies. Regardless of the tool used, which should be included in the methods with a brief description, the authors should clearly describe how data were synthesised with regard to assessing the strength of evidence, stating how RoB assessment was used and what other parameters/limitations were considered. Excluding the GRADE tool, which has clear instructions on how specific intrastudy (RoB)

and interstudy (inconsistency, imprecision, indirectness, publication bias) limitations should be handled, the authors should clarify: (1) whether study quality was determined based on RoB only or whether other parameters were taken into account (eg, precision, external validity) with a justification; (2) how consistency was defined (ie, based on direction of effect, overlap of CIs, etc). When there is some clinical heterogeneity across studies but not sufficient to preclude pooling of results, this should be accounted for and used as part of the strength of evidence assessment. More specifically, authors should include a statement to show that they have considered/assessed for potential differences in included populations, interventions and outcome measures used across studies and how they think this might affect the strength of the evidence of their results. Finally, a statement in the methods should be included to acknowledge the risk of publication bias and whether it was assessed formally. A funnel plot should be constructed and inspected/accompanied with the appropriate statistical tests (ie, Egger's test) where the results of a considerable number of studies (usually 10 or more) were pooled. For SRs including one outcome measure only, item 11 was not scored.

#### STATISTICAL ANALYSIS: SUBGROUP ANALYSIS

A mean score was calculated for all appraised SRs. Subgroup analyses included comparisons of mean scores of reviews published in a higher versus a lower impact factor (IF) journal (mean IF of included journals used as cut-off). The latest journal IFs were obtained from the journal's official website. For journals without an IF, a score of 0 was given. Finally, the mean score of the first 10 published SRs was compared with that of the last 10 published. All comparisons were performed with two-tailed independent samples t-tests in Graphpad Prism V.8. Significance levels were set at  $p < 0.05$ .

#### RESULTS

A total of 57 articles published in 38 journals were included in our review.<sup>11–67</sup> Publication date ranged from 2006 to 2020, journal IFs ranged from 0 to 59.102 (mean 4.7, median 2.55). The journal with the most published SRs of RCTs on tendinopathy interventions was the *British Journal of Sports Medicine* (n=6) and the most commonly investigated tendinopathy in isolation was shoulder (calcific and non-calcific; n=24), followed by lateral elbow (n=10) and patellar (n=8). A total of 10 reviews assessed tendinopathy as a generic condition and assessed the same intervention(s) on more than one location of tendinopathy and two reviews included RCTs of 'lower limb tendinopathies'. The most frequently assessed intervention in isolation was extracorporeal shock-wave therapy (n=10), followed by platelet-rich plasma (n=7), however, several SRs included more than one 'related' intervention (ie, 'non-surgical therapies', 'conservative treatments', 'injection therapies', etc).



Online supplemental table 1 provides an overview of the included SRs with relevant characteristics and a summary of their key findings.<sup>11–67</sup> Table 1 shows our appraisal tool and the results of the appraisal for each SR with overall scores both for each SR and for each item (explanation of the items in table caption).<sup>11–67</sup>

The most commonly used RoB assessment tools in the included SRs were the Cochrane Collaboration tool<sup>3 67</sup> (n=30) and the PEDro scale<sup>68 69</sup> (n=14). Of the 52 SRs that conducted an RoB assessment, only 23 (44%) explained how that would be used in data synthesis and 27 (52% reviews did not use their RoB assessment results in data synthesis at all. Where assessment of strength of evidence was predefined and applied, the most frequently used tools were the Cochrane Back Review Group tool<sup>4</sup> (n=10) and the GRADE tool<sup>5</sup> (n=6). Other tools for assessing strength of evidence used included the National Health and Medical Research Council (NHMRC) tool<sup>70</sup> and that devised by the Cochrane musculoskeletal group.<sup>71</sup> More than half of the included reviews (n=30; 53%) did not use or even mention quality/strength/level of evidence in their manuscript, while another nine included an arbitrary statement of a level of evidence in their discussion or conclusion without explaining the process of determining this anywhere in the manuscript. Of the remaining 18 reviews which described the use of a tool to assess the strength of evidence in their methods, only 12 of them used it appropriately according to the criteria/instructions of the tool (and to a lesser extent the judgement of the authors of the present review).

The items of our tool with the lowest overall score were item 2 (2/52; 4%) and item 11 (8/52; 15%), which both relate to whether assessments (RoB and strength of evidence) were conducted on an outcome measure and not study level. Those with the highest were item 6 (65/68; 95%) and item 4 (45/57; 79%).

The mean score of the included SRs on our appraisal tool was 46.5% (range 0%–100%). Subgroup analyses showed that those published in higher IF journals (>4.7; n=10) had a significantly higher mean score than those (n=47) in lower IF journals (68.3% vs 41.8%, difference 26.5%±8.8%, p=0.004) (figure 2A). The first 10 published SRs (2006–2012) had a slightly lower mean score (48.3%) than the 10 most recently published ones (2019–2020; 56.6%) with a statistically insignificant difference (8.3%±13.7%, p=0.54; figure 2B).

## DISCUSSION

Having critically analysed several published SRs of RCTs in the management of tendinopathy, we found that the results of more than half of the included SRs were presented without an accompanying strength of evidence assessment, which is as important as the results themselves for guiding clinical practice. Additionally, a substantial proportion of results

from the included reviews were presented with levels of evidence assigned to them arbitrarily without described and reproducible methods. Our findings imply that the majority of SR authors may not be familiar with conducting a strength of evidence assessment and its practical implementation. The other interesting finding was that half of the reviews that performed an RoB assessment did not make any use of it; we assume this is because they either do not understand the purpose of RoB assessment and they only conducted an RoB assessment knowing that it is a prerequisite for an SR, or they left the interpretation of the link between their RoB assessment and the undetermined strength of evidence open to the reader.

As an example of the disparity in assessing the strength of evidence, a recent SR of SRs on the clinical management of tendinopathy summarised all the available evidence concluding that eccentric exercise with or without other treatments is the best available intervention.<sup>72</sup> Unlike our appraisal which yielded poor overall results, their assessment of quality of the included SRs was overall high. This is predominantly because the appraisal tool they used (AMSTAR) does not take into account the importance of determining the strength of evidence for the findings. Additionally, having used very similar criteria to our review, performing searches in many more databases, and even having included some SRs with a mixture of randomised and non-randomised studies, they only included 25 SRs as opposed to 57 included in ours.

The widely used PRISMA checklist is regarded as the gold standard for the conduct and reporting of SRs.<sup>8</sup> Its 27 items cover all sections of an SR, including the title (1 item), abstract (1 item), introduction (2 items), methods (12 items), results (7 items), discussion (three items) and funding (1 item). The purpose of the checklist, as its name suggests, is to increase transparency in the reporting of SR and meta-analyses, however, the conduct of some of its items remains very subjective. This is especially true for the assessment of strength of evidence, which, according to the checklist, should be performed for each outcome measure in the discussion section, however, further guidance/instructions are not provided.<sup>8</sup> We argue that details of the process of strength of evidence assessment should be included in the methods section of an SR to increase transparency and reproducibility.

## Recommendations

In the following paragraphs, we provide recommendations for what we believe are the most subjective and poorly understood aspects of an SR, aiming to increase transparency, reproducibility and hence the quality of the SR and the confidence in its findings. According to the National Institute for Health Research, authors should only identify their article as an SR ‘when the account of the search, appraisal and

**Table 1** Appraisal results of each study using our tool with total scores and percentages both for each study and for each tool item

Study	Risk of bias (RoB) assessment		Pooling of results												Strength of evidence assessment				Total score	%
	1 (0-2)	2 (0-1)	3 (0-2)	4 (0-1)	5 (0-2)	6 (0-2)	7 (0-2)	8 (0-2)	9 (0-2)	10 (0-2)	11 (0-1)	12 (0-1)	11 (0-1)	12 (0-1)						
Arirachakaran <i>et al</i> <sup>11</sup>	0	0	2	1	-	2	0	0	0	0	2	0	0	0	0	0	7/18	39		
Arirachakaran <i>et al</i> <sup>12</sup>	0	0	2	1	0	2	0	0	0	0	2	0	0	0	0	0	7/20	35		
Balasubramaniam <sup>13</sup>	2	0	2	1	2	-	0	2	2	0	0	0	0	0	0	0	11/18	61		
Bannuru <i>et al</i> <sup>14</sup>	2	0	2	1	2	-	0	0	0	0	0	0	0	0	0	0	7/18	39		
Bjordan <i>et al</i> <sup>15</sup>	2	0	2	1	-	2	0	0	0	2	0	0	0	0	0	0	9/18	50		
Boudreault <i>et al</i> <sup>16</sup>	0	0	2	0	-	2	0	0	1	2	0	0	0	0	0	0	7/18	39		
Buchbinder <i>et al</i> <sup>17</sup>	2	0	2	1	2	2	2	2	2	0	1	1	1	1	1	17/20	85			
Catapano <i>et al</i> <sup>18</sup>	0	0	2	0	0	-	0	0	0	0	0	0	0	0	0	0	2/18	11		
Challoumas <i>et al</i> <sup>19</sup>	2	0	0	1	2	-	2	0	2	0	2	0	1	1	1	11/18	61			
Challoumas <i>et al</i> <sup>20</sup>	2	0	0	1	2	-	2	0	2	0	2	0	1	1	1	7/11	61			
Chen <i>et al</i> <sup>21</sup>	0	0	2	1	-	2	0	2	0	2	0	0	0	0	0	0	9/18	50		
Coombes <i>et al</i> <sup>22</sup>	2	0	2	1	2	2	2	2	2	2	2	0	0	0	0	1	17/20	85		
Dan <i>et al</i> <sup>23</sup>	2	1	2	1	2	2	2	2	2	2	2	2	1	1	1	20/20	100			
de Vos <i>et al</i> <sup>24</sup>	2	0	0	1	1	-	0	0	0	2	0	0	0	0	0	0	6/18	33		
Desjardins-Charbonneau <i>et al</i> <sup>25</sup>	0	0	2	1	0	2	0	2	0	2	0	2	0	0	0	0	9/20	45		
Desjardins-Charbonneau <i>et al</i> <sup>26</sup>	0	0	2	0	0	2	0	2	1	2	0	0	0	0	0	0	9/20	45		
Desmeules <i>et al</i> <sup>27</sup>	0	0	0	1	0	-	0	0	1	0	0	0	0	0	0	0	2/18	11		
Desmeules <i>et al</i> <sup>28</sup>	0	0	0	0	0	-	0	0	0	0	0	0	0	0	0	0	0/18	0		
Desmeules <i>et al</i> <sup>29</sup>	0	0	2	0	0	2	0	0	1	0	0	0	0	0	0	0	5/20	25		
Dong <i>et al</i> <sup>30</sup>	2	-	2	1	-	2	1	0	0	0	0	0	-	-	0	0	8/16	50		
Dupleix and Charalambous <sup>31</sup>	0	-	2	1	-	2	1	0	0	0	0	0	-	-	0	0	6/16	38		
Fitzpatrick <i>et al</i> <sup>32</sup>	1	-	2	1	-	2	2	2	1	0	0	0	-	-	0	0	11/16	69		
Haslerud <i>et al</i> <sup>33</sup>	2	0	2	1	-	2	0	2	1	0	0	0	0	0	0	0	10/18	56		
Heales <i>et al</i> <sup>34</sup>	2	0	2	1	2	-	0	2	2	0	0	1	1	1	1	13/18	72			
Huisstede <i>et al</i> <sup>35</sup>	2	0	0	1	0	-	2	0	2	0	0	0	0	0	0	0	7/18	39		
Ioppolo <i>et al</i> <sup>36</sup>	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1/20	5		
Krey <i>et al</i> <sup>37</sup>	0	0	0	0	0	-	0	0	0	0	0	0	0	0	0	0	0/18	0		
Lafrance <i>et al</i> <sup>38</sup>	0	0	2	1	0	-	1	2	1	0	0	0	0	0	0	0	9/18	50		
Larsson <i>et al</i> <sup>39</sup>	1	0	0	0	0	-	0	0	2	0	0	0	0	0	0	0	3/18	17		

Continued



Table 1 Continued

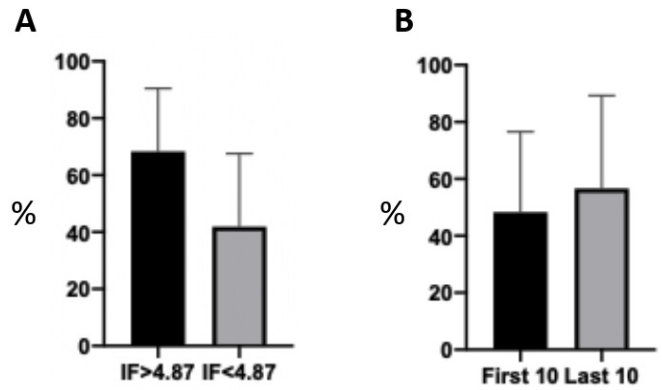
Study	Risk of bias (RoB) assessment		Pooling of results					Strength of evidence assessment					Total score	%
	1 (0-2)	2 (0-1)	3 (0-2)	4 (0-1)	5 (0-2)	6 (0-2)	7 (0-2)	8 (0-2)	9 (0-2)	10 (0-2)	11 (0-1)	12 (0-1)		
Lee <i>et al</i> <sup>40</sup>	1	0	0	0	0	-	2	0	2	0	0	1	6/18	33
Li <i>et al</i> <sup>77</sup>	0	0	2	0	-	2	1	2	0	0	0	0	7/18	39
Liao <i>et al</i> <sup>41</sup>	0	0	2	1	-	2	2	0	0	2	0	0	9/18	50
Lin <i>et al</i> <sup>42</sup>	2	0	2	1	-	2	2	2	0	2	0	0	13/18	72
Lin <i>et al</i> <sup>43</sup>	2	1	2	1	-	2	2	2	0	2	0	0	14/18	78
Lin <i>et al</i> <sup>44</sup>	1	0	2	1	-	2	2	0	0	2	0	0	10/18	56
Littlewood <i>et al</i> <sup>45</sup>	2	0	0	1	0	-	1	0	2	0	0	0	6/18	33
Louwerens <i>et al</i> <sup>46</sup>	1	0	2	1	0	2	1	2	2	0	1	1	14/20	70
Maffulli <i>et al</i> <sup>47</sup>	0	0	0	0	0	-	0	0	0	0	0	0	0/18	0
Magnussen <i>et al</i> <sup>48</sup>	0	0	0	1	0	-	0	0	0	0	0	0	1/18	6
Martimbianco <i>et al</i> <sup>49</sup>	2	0	2	1	-	2	1	2	2	2	1	1	16/18	89
Mendonça <i>et al</i> <sup>50</sup>	2	0	2	1	-	2	2	0	2	2	1	1	15/18	83
Miller <i>et al</i> <sup>51</sup>	0	0	2	1	-	2	2	0	1	2	0	0	10/18	56
Mohamadi <i>et al</i> <sup>52</sup>	2	-	2	1	2	2	2	2	0	2	-	0	15/18	83
Murphy <i>et al</i> <sup>53</sup>	2	-	2	1	-	2	0	2	2	2	-	1	14/16	88
Nogueira and Moura <sup>54</sup>	0	0	0	0	0	-	0	0	0	0	0	0	0/18	0
Ortega-Castillo and Medina-Portuques <sup>55</sup>	2	0	2	1	2	-	0	0	2	0	0	0	9/18	50
Sussmilch-Leitch <i>et al</i> <sup>56</sup>	2	0	2	1	2	2	1	2	0	0	0	0	13/20	65
Tsikopoulos <i>et al</i> <sup>57</sup>	0	0	2	1	2	-	2	2	0	2	0	0	11/18	61
Toliopoulos <i>et al</i> <sup>58</sup>	0	0	0	0	0	-	0	0	1	0	0	0	1/18	6
Vander Doelen and Jelley <sup>59</sup>	0	0	0	1	0	-	0	0	0	0	0	0	1/18	6
Verstraelen <i>et al</i> <sup>60</sup>	0	0	2	1	2	2	2	2	0	2	0	0	13/20	65
Wasilewski and Kotsko <sup>61</sup>	0	0	2	1	2	-	0	0	0	0	0	0	5/18	28
Woodley <i>et al</i> <sup>62</sup>	2	0	2	1	2	2	1	2	2	0	0	1	15/20	75
Wu <i>et al</i> <sup>63</sup>	0	0	2	1	-	2	2	2	0	2	0	0	11/18	61
Xiong <i>et al</i> <sup>64</sup>	2	0	2	1	-	2	0	0	0	0	0	0	7/18	39
Yan <i>et al</i> <sup>65</sup>	0	0	2	1	-	1	1	0	0	0	0	0	5/18	28
Zhang <i>et al</i> <sup>66</sup>	0	0	2	1	2	2	0	2	0	2	0	0	11/20	55
Overall mean score	-	-	-	-	-	-	-	-	-	-	-	-	-	46.5

Continued

Table 1 Continued

Study	Risk of bias (RoB) assessment			Pooling of results						Strength of evidence assessment				
	1 (0-2)	2 (0-1)	3 (0-2)	4 (0-1)	5 (0-2)	6 (0-2)	7 (0-2)	8 (0-2)	9 (0-2)	10 (0-2)	11 (0-1)	12 (0-1)	Total score	%
Overall Total for item	53/114	2/52	82/114	45/57	33/74	65/68	46/114	48/114	45/114	44/114	8/52	12/57		
%	47%	4%	72%	79%	45%	96%	40%	42%	39%	39%	15%	21%		

Item 1—was RoB assessment used in data synthesis and was it prespecified/explained in methods? ('0' for neither, '2' for both, '1' if used but not prespecified) —(PRISMA items 12, 19). Item 2—was RoB assessment performed on an outcome measure level? If not, do the authors state why in methods? ('1' if either fulfilled, '0' if neither) —(PRISMA items 12, 19). Item 3—are principally summary measures predefined in methods and used in results? ('0' for neither, '2' for both, '1' if used but not prespecified) —(PRISMA item 13). Item 4—was the outcome measures of the SR predefined in methods? ('0' for no, '1' for yes) —(PRISMA item 11). Item 5—is qualitative synthesis appropriate and was it prespecified/explained in methods? ('0' for neither or if only narrative description without pooling of results, '2' for both, '1' if appropriate but not prespecified or if prespecified but not done) —(PRISMA items 13, 14). Item 6—is quantitative synthesis appropriate and was it prespecified/explained in methods? ('0' for neither, '2' for both, '1' if appropriate but not prespecified) —(PRISMA items 20, 21). Item 7—was follow-up time points for data analyses predefined and used? ('0' for neither, '2' for both, '1' if used but not prespecified) —(Cochrane Back Review Group). Item 8—was missing data dealt with appropriately and was this prespecified in methods? ('0' for neither, '2' for both, '1' if appropriate but not prespecified) —(Cochrane Back Review Group). Item 9—was assessment of the strength of the evidence used and was it prespecified/explained in methods? ('0' for neither, '2' for both, '1' if used but not prespecified) —(PRISMA item 24). Item 10—was the risk of publication bias acknowledged in methods and was it assessed where appropriate? ('0' for neither, '2' for both, '1' if done but not prespecified) —(PRISMA items 15, 25). Item 11—was the strength of the evidence assessed on an outcome measure level? ('1' for yes, '0' for no or if assessment of strength of evidence not performed) —(GRADE). Item 12—was the strength of the evidence assessed appropriately according to the prespecified tool? ('1' for yes, '0' for no or if not performed/prespecified) (Strength of evidence assessment tools used).  
 GRADE, Grading of Recommendations Assessment, Development and Evaluation; PRISMA, Preferred Reporting Items in Systematic Reviews and Meta-Analyses; SR, systematic review.



**Figure 2** Subgroup analyses comparing the mean scores of reviews published in a higher versus a lower impact factor (IF) journals. Those published in higher if journals (>4.7; n=10) had a significantly higher mean score than those (n=47) in lower if journals (68.3% vs 41.8%, difference 26.5%±8.8%, p=0.004) (A). The first 10 published SRs (2006–2012) had a slightly lower mean score (48.3%) than the 10 most recently published ones (2019–2020; 56.6%) with a statistically insignificant difference (8.3%±13.7%, p=0.54; B). SRs, systematic reviews.

synthesis methods would, in theory, permit the replication of the review by others'.<sup>7</sup> All recommendations arise from key documents guiding the conduct of SRs: the PRISMA checklist, the Cochrane handbook, the GRADE handbook and the Cochrane Collaboration Back Review Group guidance.<sup>4 8 67 73</sup>

**RoB assessment**

We highlight once again that an RoB assessment is the most important part of but not exactly synonymous to study quality assessment. Where the strength of evidence assessment tool used relies on study quality (eg, Cochrane Back Review Group tool), the authors should clarify if they only use RoB assessment as an extension of study quality. Finally, RoB assessment should be performed on an outcome measure level, not a study level, especially when the study includes both patient-reported outcomes and outcomes directly measured by study faculty.

We recommend the use of the revised RoB assessment tool recently published by the Cochrane Collaboration (RoB 2) as it is more standardised and reproducible with clear instructions on how it should be used and less subjectivity.<sup>74</sup> Most importantly, the creators include instructions for determination of overall RoB for each study which will substantially increase its correct implementation in data syntheses in SRs.<sup>74</sup> Alternatively, the RCT methodology checklist published by the Scottish Intercollegiate Guideline Network<sup>75</sup> could also be used, especially where the Cochrane Collaboration Back Review Group tool is used for strength of evidence assessment (see below). Importantly, the checklist finishes with an 'overall assessment of the study', which is the equivalent of RoB 2's 'overall RoB' and can be used directly for strength of evidence assessment.



## Pooling of results

All SRs should include pooling of results to make conclusions about an overall treatment effect where clinical heterogeneity allows.<sup>8</sup>

In the absence of significant clinical heterogeneity, results should be pooled quantitatively where there are two or more studies assessing the same interventions, ideally with pairwise meta-analyses that improve precision.<sup>8,67</sup> In the context of tendinopathy interventions, unless the settings, participants and interventions are almost identical (clinical homogeneity), we suggest the use of a random-effects and not a fixed-effects model for meta-analyses.<sup>67</sup> Alternatively, if the authors of the review believe that the included studies are largely homogenous clinically, the choice of the model should be made on the basis of statistical heterogeneity, which needs to be quantified with a statistical test (eg, fixed-effects for  $I^2 < 50\%$ , random-effects for  $I^2 > 50\%$ ).<sup>76</sup> Forest plots should be presented for each comparison with their accompanying MD/SMD (continuous outcomes) or OR/RR (dichotomous outcomes) and 95% CI, p value and heterogeneity test (eg,  $I^2$  statistic).<sup>8,67</sup> In the presence of significant statistical heterogeneity (eg,  $> 50\%$ ), attempts should be made through sensitivity analyses for a single study that may be responsible for the high heterogeneity to be identified and removed from the comparison before the analysis is rerun.<sup>67,76</sup> No more than one study should be removed. Where the  $I^2$  test is substantial ( $> 75\%$ ) despite sensitivity analyses, the meta-analysis should be abandoned.<sup>76</sup>

For qualitative reviews, where numerical values of treatment effects cannot be obtained and pooled quantitatively, pooling should be performed based on direction of effect for studies assessing the same interventions, even in the presence of some clinical heterogeneity, which should be recognised and accounted for in the strength of evidence assessment.<sup>4</sup> The statistical significance test result reported by the authors of the RCT for intergroup differences should be accounted for when pooling results (ie, a statistically insignificant difference between the two groups will count as ‘no difference’). A table with pooled results and accompanying strengths of evidence should be used to make the pooling clearer to the reader (eg, table 1).

Data should only be combined for similar, predefined follow-up time periods, that is, short term, mid term and long-term.<sup>4</sup> These periods should ideally be subdivided especially when their range is large and when there are sufficient data for pooling. For instance, for a predefined

‘short-term follow-up’ at 0–12 weeks, this can be subdivided into early short term (eg, 2–6 weeks) and late short term (7–12 weeks).

## Strength of evidence assessment

The two tools most frequently used in our included SRs (Cochrane Back Review Group and GRADE) differ substantially and we advocate the preferential use of one over the other in specific situations.<sup>4,73</sup> Unlike the Cochrane Back Review Group tool, the intention of which is to allocate a ‘level’ (strength) of evidence to data pooled qualitatively, the GRADE tool is very comprehensive, it considers limitations across studies, not only within studies and is very transparent.<sup>4,73</sup> The main difference of the two tools lies within the fact that the Cochrane Back Review Group tool relies on quality assessment of each study; the GRADE tool advocates assessing the quality of the body of the pooled evidence, not individual studies. The former is primarily based on direction of effect (ie, qualitative synthesis of results) and the latter on quantitative analyses.

Considering the nature and the intention of these two tools, we recommend the use of the GRADE tool for all SRs where comparisons of interventions include any type of quantitative analyses (with or without a meta-analysis) and the Cochrane Back Review Group where only direction of effect is used for pooling of results (ie, qualitative analysis only). However, when the Cochrane Back Review Group tool is used, in addition to RoB assessment which should be used to determine the level of evidence for each outcome measure, we recommend using imprecision, indirectness of evidence and publication bias (where appropriate) to downgrade the level of evidence in a similar fashion to the GRADE tool. Besides, van Tulder *et al* state in their original publication that in addition to methodological quality of the original studies, participants, interventions and outcomes should also be taken into account when levels of evidence are attributed to qualitative analyses.<sup>4</sup> Table 2 and table 3 can be used to increase transparency of the methods used for assessing the strength of evidence. Alternatively, the ‘summary of findings’ table suggested by GRADE can be used, separately for each outcome measure and follow-up time period with explanations for each downgrading of the evidence.<sup>73</sup>

Neither the GRADE tool nor the Cochrane Collaboration BackReviewGroup tool come without disadvantages.<sup>45</sup> Despite GRADE’s comprehensiveness, which includes a detailed

**Table 2** Suggested pooling of results for qualitative analyses (example)

Comparison	Follow-up	Study	Outcome measure 1	Outcome measure 2	Outcome measure 3
Intervention 1 versus intervention 2	Short/mid/long term	Author (year)	↓	↓	↓
		Author (year)	↓	↔	↓
		Author (year)	↓	↑	–
Overall intervention 1 versus intervention 2 (evidence level)			↓ (moderate)	– (conflicting)	↓ (limited)

**Table 3** Suggested description of strength of evidence assessment

Comparison	Outcome measure	Follow-up	Overall RoB	Inconsistency	Imprecision	Indirectness	Publication bias	Strength of evidence
Intervention <sub>1</sub> vs Intervention <sub>2</sub>	Outcome measure 1	Short term	High/low/unclear	High/low risk	High/low risk	High/low risk	High/low risk	High/moderate/low (very low, limited, conflicting)
		Mid term	High/low/unclear	High/low risk	High/low risk	High/low risk	High/low risk	High/moderate/low (very low, limited, conflicting)
		Long term	High/low/unclear	High/low risk	High/low risk	High/low risk	High/low risk	High/moderate/low (very low, limited, conflicting)
	Outcome measure 2	Short term	High/low/unclear	High/low risk	High/low risk	High/low risk	High/low risk	High/moderate/low (very low, limited, conflicting)
		Mid term	High/low/unclear	High/low risk	High/low risk	High/low risk	High/low risk	High/moderate/low (very low, limited, conflicting)
		Long term	High/low/unclear	High/low risk	High/low risk	High/low risk	High/low risk	High/moderate/low (very low, limited, conflicting)
	Outcome measure 3	Short term	High/low/unclear	High/low risk	High/low risk	High/low risk	High/low risk	High/moderate/low (very low, limited, conflicting)
		Mid term	High/low/unclear	High/low risk	High/low risk	High/low risk	High/low risk	High/moderate/low (very low, limited, conflicting)
		Long term	High/low/unclear	High/low risk	High/low risk	High/low risk	High/low risk	High/moderate/low (very low, limited, conflicting)

RoB, risk of bias.

handbook providing explanation and instructions of all the steps, its conduct requires decisions that are based on the assessor's judgement, for example, determining whether the overall RoB for an outcome measure is high or low when pooling studies with overall high and low RoB and deciding whether the magnitude of inconsistency and indirectness is significant enough to justify downgrading the evidence.<sup>73</sup> Nevertheless, we believe that it is the most thorough and transparent tool for strength of evidence assessment and it should be used preferentially over other tools where possible. Similarly, the Cochrane Collaboration Back Review Group tool requires subjective decisions for the determination of the strength of evidence for a result, primarily due to its lack of definition for its two main criteria; (1) consistency of results and (2) study quality. Additionally, unlike GRADE, it does not consider 'between-study' limitations, such as imprecision of results and indirectness of evidence.

An SR is no small undertaking and should not make for an easy publication. Future authors who only intend to summarise the available evidence without a comprehensive quality assessment and strength of evidence assessment should strongly consider identifying their reviews as narrative, not systematic. Admittedly, strict word counts and other restrictions from journals undoubtedly have a negative influence on the quality of the published SRs; we, therefore, invite the editors of publishing journals to either loosen restrictions for submitted articles or discuss word counts and accompanying material (tables/figures) with authors on a case-by-case basis. Publication of narrative reviews should also become an option in all publishing journals.

## CONCLUSIONS

In the present review, we have demonstrated that the majority of SRs of RCTs in tendinopathy do not assess the strength of evidence of their results and this can substantially influence their application to clinical practice. In the future, authors

implicated in producing and publishing SRs on the management of tendinopathy should become familiar with strength of evidence assessment tools and apply them appropriately in their data synthesis and presentation. Finally, relevant guidance documents should emphasise the importance of strength of evidence assessment and provide more detailed instructions for its conduct to increase consistency and transparency.

**Twitter** Neal L Millar @tendonglasgow

**Contributors** DC and NLM conceived and designed the study. DC and NLM performed analysis. All authors analysed the data. DC and NLM wrote the paper.

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

**Competing interests** None declared.

**Patient consent for publication** Not required.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution 4.0 Unported (CC BY 4.0) license, which permits others to copy, redistribute, remix, transform and build upon this work for any purpose, provided the original work is properly cited, a link to the licence is given, and indication of whether changes were made. See: <https://creativecommons.org/licenses/by/4.0/>.

## ORCID iD

Neal L Millar <http://orcid.org/0000-0001-9251-9907>

## REFERENCES

- 1 Cochrane. Cochrane consumer network. Available: <https://consumers.cochrane.org/levels-evidence> [Accessed 10 Jun 2020].
- 2 Pae C-U. Why systematic review rather than narrative review? *Psychiatry Investig* 2015;12:417-9.
- 3 Higgins JPT, Altman DG, Gøtzsche PC, et al. The Cochrane collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 2011;343:d5928.
- 4 van Tulder M, Furlan A, Bombardier C. Editorial board of the Cochrane collaboration back review group. updated method guidelines for systematic reviews in the Cochrane collaboration back review group. *Spine* 2003;28:1290-9.

- 5 Guyatt GH, Oxman AD, Vist GE, *et al.* Grade: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008;336:924–6.
- 6 Definition of levels of evidence (LOE) and overall strength of evidence (SoE). *Global Spine J* 2015;5:539.
- 7 National Institute for Health Research (NIHR). Evidence synthesis and systematic reviews. Available: <https://www.journalslibrary.nihr.ac.uk/information-for-authors/evidence-synthesis-and-systematic-reviews/> [Accessed 15 Jun 2020].
- 8 Moher D, Liberati A, Tetzlaff J, *et al.* Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 2009;6:e1000097.
- 9 Shea BJ, Reeves BC, Wells G, *et al.* AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ* 2017;358:j4008.
- 10 Horton R. Offline: the gravy train of systematic reviews. *Lancet* 2019;394:1790.
- 11 Arirachakaran A, Sukthuyat A, Sisayanarane T, *et al.* Platelet-Rich plasma versus autologous blood versus steroid injection in lateral epicondylitis: systematic review and network meta-analysis. *J Orthop Traumatol* 2016;17:101–12.
- 12 Arirachakaran A, Boonard M, Yamaphai S, *et al.* 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;27:381–90.
- 13 Balasubramaniam U, Dissanayake R, Annabell L. Efficacy of platelet-rich plasma injections in pain associated with chronic tendinopathy: a systematic review. *Phys Sportsmed* 2015;43:253–61.
- 14 Bannuru RR, Flavin NE, Vaysbrot E, *et al.* High-Energy extracorporeal shock-wave therapy for treating chronic calcific tendinitis of the shoulder: a systematic review. *Ann Intern Med* 2014;160:542–9.
- 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 J-S, *et al.* The efficacy of oral non-steroidal anti-inflammatory drugs for rotator cuff tendinopathy: a systematic review and meta-analysis. *J Rehabil Med* 2014;46:294–306.
- 17 Buchbinder R, Green SE, Youd JM, *et al.* Systematic review of the efficacy and safety of shock wave therapy for lateral elbow pain. *J Rheumatol* 2006;33:1351–63.
- 18 Catapano M, Catapano J, Borschel G, *et al.* 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;101:897–906.
- 19 Challoumas D, Clifford C, Kirwan P, *et al.* 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:e000528.
- 20 Challoumas D, Kirwan PD, Borysov D, *et al.* Topical glyceryl trinitrate for the treatment of tendinopathies: a systematic review. *Br J Sports Med* 2019;53:251–62.
- 21 Chen P-C, Wu K-T, Chou W-Y, *et al.* Comparative effectiveness of different nonsurgical treatments for patellar tendinopathy: a systematic review and network meta-analysis. *Arthroscopy* 2019;35:3117–31.
- 22 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;376:1751–67.
- 23 Dan M, Phillips A, Johnston RV. Surgery for patellar tendinopathy (jumper's knee). *Cochrane Database Syst Rev* 2019;88.
- 24 de Vos R-J, Windt J, Weir A. Strong evidence against platelet-rich plasma injections for chronic lateral epicondylar tendinopathy: a systematic review. *Br J Sports Med* 2014;48:952–6.
- 25 Desjardins-Charbonneau A, Roy J-S, Dionne CE, *et al.* The efficacy of taping for rotator cuff tendinopathy: a systematic review and meta-analysis. *Int J Sports Phys Ther* 2015;10:420–33.
- 26 Desjardins-Charbonneau A, Roy J-S, Dionne CE, *et al.* The efficacy of manual therapy for rotator cuff tendinopathy: a systematic review and meta-analysis. *J Orthop Sports Phys Ther* 2015;45:330–50.
- 27 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;58:389–403.
- 28 Desmeules F, Boudreault J, Roy J-S, *et al.* Efficacy of transcutaneous electrical nerve stimulation for rotator cuff tendinopathy: a systematic review. *Physiotherapy* 2016;102:41–9.
- 29 Desmeules F, Boudreault J, Roy J-S, *et al.* The efficacy of the therapeutic ultrasound for rotator cuff tendinopathy: a systematic review and meta-analysis. *Phys Ther Sport* 2015;16:276–84.
- 30 Dong W, Goost H, Lin X-B, *et al.* Injection therapies for lateral epicondylalgia: a systematic review and Bayesian network meta-analysis. *Br J Sports Med* 2016;50:900–8.
- 31 Dupley L, Charalambous CP. Platelet-Rich plasma injections as a treatment for refractory patellar Tendinosis: a meta-analysis of randomised trials. *Knee Surg Relat Res* 2017;29:165–71.
- 32 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;45:226–33.
- 33 Haslerud S, Magnussen LH, Joensen J, *et al.* The efficacy of Low-Level laser therapy for shoulder tendinopathy: a systematic review and Meta-Analysis of randomized controlled trials. *Physiother Res Int* 2015;20:108–25.
- 34 Heales LJ, McClintock SR, Maynard S, *et al.* Evaluating the immediate effect of forearm and wrist orthoses on pain and function in individuals with lateral elbow tendinopathy: a systematic review. *Musculoskelet Sci Pract* 2020;47:102147.
- 35 Huisstede BMA, Gebremariam L, van der Sande R, *et al.* Evidence for effectiveness of Extracorporeal Shock-Wave Therapy (ESWT) to treat calcific and non-calcific rotator cuff tendinosis--a systematic review. *Man Ther* 2011;16:419–33.
- 36 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;94:1699–706.
- 37 Krey D, Borchers J, McCamey K. Tendon needling for treatment of tendinopathy: a systematic review. *Phys Sportsmed* 2015;43:80–6.
- 38 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:e000506.
- 39 Larsson MEH, Käll I, Nilsson-Helander K. Treatment of patellar tendinopathy--a systematic review of randomized controlled trials. *Knee Surg Sports Traumatol Arthrosc* 2012;20:1632–46.
- 40 Lee S-Y, 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;20:845–54.
- 41 Liao C-D, Tsao J-Y, Chen H-C, *et al.* Efficacy of extracorporeal shock wave therapy for lower-limb tendinopathy: a meta-analysis of randomized controlled trials. *Am J Phys Med Rehabil* 2018;97:605–19.
- 42 Lin M-T, Wei K-C, Wu C-H. Effectiveness of platelet-rich plasma injection in rotator cuff tendinopathy: a systematic review and meta-analysis of randomized controlled trials. *Diagnostics* 2020;10:189.
- 43 Lin M-T, Chiang C-F, Wu C-H, *et al.* 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;100:336–49.
- 44 Lin Y-C, Wu W-T, Hsu Y-C, *et al.* Comparative effectiveness of botulinum toxin versus non-surgical treatments for treating lateral epicondylitis: a systematic review and meta-analysis. *Clin Rehabil* 2018;32:131–45.
- 45 Littlewood C, Ashton J, Chance-Larsen K, *et al.* Exercise for rotator cuff tendinopathy: a systematic review. *Physiotherapy* 2012;98:101–9.
- 46 Louwerens JKG, Siersevelt IN, van Noort A, *et al.* 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;23:1240–9.
- 47 Maffulli N, Papalia R, D'Adamo S, *et al.* Pharmacological interventions for the treatment of Achilles tendinopathy: a systematic review of randomized controlled trials. *Br Med Bull* 2015;113:101–15.
- 48 Magnussen RA, Dunn WR, Thomson AB. Nonoperative treatment of midportion Achilles tendinopathy: a systematic review. *Clin J Sport Med* 2009;19:54–64.
- 49 Martimbianco ALC, Ferreira RES, Latorraca CdeOC, *et al.* Photobiomodulation with low-level laser therapy for treating Achilles tendinopathy: a systematic review and meta-analysis. *Clin Rehabil* 2020;34:713–22.
- 50 Mendonça LDM, Leite HR, Zwerter J, *et al.* 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;54:87–93.
- 51 Miller LE, Parrish WR, Roides B, *et al.* 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:e000237.
- 52 Mohamadi A, Chan JJ, Claessen FMAP, *et al.* Corticosteroid injections give small and transient pain relief in rotator cuff Tendinosis: a meta-analysis. *Clin Orthop Relat Res* 2017;475:232–43.
  - 53 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;53:1070–7.
  - 54 Nogueira AC, Júnior MdeJM. The effects of laser treatment in tendinopathy: a systematic review. *Acta Ortop Bras* 2015;23:47–9.
  - 55 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;19:438–53.
  - 56 Sussmilch-Leitch SP, Collins NJ, Bialocerkowski AE, *et al.* Physical therapies for Achilles tendinopathy: systematic review and meta-analysis. *J Foot Ankle Res* 2012;5:15.
  - 57 Tsikopoulos K, Tsikopoulos A, Natsis K. Autologous whole blood or corticosteroid injections for the treatment of epicondylar and plantar fasciopathy? A systematic review and meta-analysis of randomized controlled trials. *Phys Ther Sport* 2016;22:114–22.
  - 58 Toliopoulos P, Desmeules F, Boudreault J, *et al.* Efficacy of surgery for rotator cuff tendinopathy: a systematic review. *Clin Rheumatol* 2014;33:1373–83.
  - 59 Vander Doelen T, Jelley W. Non-Surgical treatment of Patellar tendinopathy: a systematic review of randomized controlled trials. *J Sci Med Sport* 2020;23:118–24.
  - 60 Verstraelen FU, In den Kleef NJHM, Jansen L, *et al.* High-Energy versus low-energy extracorporeal shock wave therapy for calcifying tendinitis of the shoulder: which is superior? A meta-analysis. *Clin Orthop Relat Res* 2014;472:2816–25.
  - 61 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;42:409–21.
  - 62 Woodley BL, Newsham-West RJ, Baxter GD. Chronic tendinopathy: effectiveness of eccentric exercise. *Br J Sports Med* 2007;41:188–99.
  - 63 Wu Y-C, Tsai W-C, Tu Y-K, *et al.* 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;98:1678–92.
  - 64 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;47:284–9.
  - 65 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;14:248.
  - 66 Zhang T, Duan Y, Chen J, *et al.* Efficacy of ultrasound-guided percutaneous lavage for rotator cuff calcific tendinopathy: a systematic review and meta-analysis. *Medicine* 2019;98:e15552.
  - 67 Cochrane. Cochrane Handbook for systematic reviews of interventions. online. Available: <https://training.cochrane.org/handbook/current> [Accessed 10 Jun 2020].
  - 68 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;51:1235–41.
  - 69 PEDRO Scale. Physiotherapy evidence database. online. Available: <http://www.pedro.org.au/English/downloads/pedro-scale/> [Accessed 9 Jun 2020].
  - 70 National Health and Medical Research Council. Australian government: NHMRC additional levels of evidence and grades for recommendations for developers of guidelines, 2009. Available: <http://www.nhmrc.gov.au/guidelines/developers.htm#add> [Accessed 5 Jun 2020].
  - 71 Tugwell P, Shea B, Boers M. *Evidence-Based rheumatology*. London: BMJ Books, 2003.
  - 72 Irby A, Gutierrez J, Chamberlin C, *et al.* Clinical management of tendinopathy: a systematic review of systematic reviews evaluating the effectiveness of tendinopathy treatments. *Scand J Med Sci Sports* 2020;30:1810–26.
  - 73 GRADE Handbook. Handbook for grading the quality of evidence and the strength of recommendations using the grade approach. Available: <https://gdt.gradeapro.org/app/handbook/handbook.html> [Accessed 12 Jun 2020].
  - 74 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.
  - 75 Scottish Intercollegiate Guideline Network (SIGN). Sign 50: checklists. Edinburgh: sign. online. Available: <https://www.sign.ac.uk/what-we-do/methodology/checklists/> [Accessed 17 Dec 2020].
  - 76 Ryan R. Cochrane Consumers and Communication Review Group. 'Heterogeneity and subgroup analyses in Cochrane Consumers and Communication Group reviews: planning the analysis at protocol stage, 2016. Available: <http://cccrg.cochrane.org> [Accessed 15 Jun 2020].
  - 77 Li A, Wang H, Yu Z, *et al.* Platelet-Rich plasma vs corticosteroids for elbow epicondylitis: a systematic review and meta-analysis. *Medicine* 2019;98:e18358.



## Supplementary Figure 1: Useful definitions

**Study quality:** the extent to which its design, conduct, analysis and presentation were appropriate to answer its research question. Consists of internal validity, external validity and precision

**Quality/strength of evidence:** one's confidence in the estimates of effect. This is usually "high", "moderate", "low", "very low", "limited" or "conflicting" depending on the grading tool used

**Internal validity of a study:** freedom from methodological bias (extent to which confounding factors were minimised); i.e. did the study use a sound methodology?

**External validity of a study:** the extent to which the results of the study are applicable and generalisable to the target population in clinical practice. This depends on the population, interventions and outcome measures used in studies and how they relate to those in clinical practice.

**Precision:** freedom from random error, i.e. the likelihood of the results of the studies to represent the true effect of the tested intervention. Generally, the smaller the population and the fewer the events in a study, the greater the imprecision (reflected by a wider confidence interval). Similarly, in the context of a systematic review/meta-analysis, the wider the confidence interval of the estimate of the treatment effect the higher the imprecision of the pooled evidence.

**Directness of evidence:** evidence arising from direct comparisons of interventions, delivered to the populations in which we are interested, and measures the outcomes important to patients. The four sources of indirectness are: differences in populations, differences in interventions, differences in outcome measures (surrogate outcomes) and indirect comparisons. Indirectness of evidence is synonymous to clinical

heterogeneity when it is assessed on an “across studies” (“inter-study”) level and external validity when on a “within studies” (“intra-study”) level.

**Consistency of results:** the extent to which all included studies have the same underlying effect. In the context of a systematic review/meta-analysis, inconsistency of results across studies is synonymous to statistical heterogeneity, which is usually quantified with the  $I^2$  statistic. The  $I^2$  statistic quantifies the proportion of the variation in point estimates due to among-study differences.

Authors	Journal	Tendinopath	Numbe	Intervention	Outcome	Risk of	Strength of	Summary of	Meta-
---------	---------	-------------	-------	--------------	---------	---------	-------------	------------	-------

	(Impact Factor)	y	r of included studies	Assessed	Measures	Bias Assessment Tool	Evidence Assessment Tool	Findings	analysis
<b>Arirachakaran et al. (2017)<sup>11</sup></b>	J Orthop Traumatol (1.273)	Lateral Elbow	10	PRP, Autologous blood, corticosteroid injection	Pain, function, pain pressure threshold	Cochrane	None	PRP van improve pain and fewer complications. Autologous blood can improve pain, function and pain pressure thresholds but has higher complication rates.	Yes
<b>Arirachakaran et al. (2017)<sup>12</sup></b>	Eur J Orthop Surg Traumatol (1.560)	Shoulder calcific	7	ESWT, US-guided lavage, corticosteroid injection and combined treatment	Pain, function	Cochrane	None	US-guided lavage is the treatment of choice	Yes
<b>Balasubramanian et al. (2015)<sup>13</sup></b>	Phys Sportsmed (1.874)	All	9	PRP	Pain, disability	Furlan et al. (2009)	Cochrane Back Review Group	PRP looks promising but more research needed	No
<b>Bannuru et al. (2014)<sup>14</sup></b>	Ann Intern Med	Shoulder calcific	28	ESWT	Pain, function,	Cochrane	None	High-energy ESWT is	No



	(19.315)				resolution of calcifications			effective at improving pain and function	
<b>Bjordal et al. (2008)<sup>15</sup></b>	BMC Musculoskeletal Disord (2.050)	Lateral Elbow	18	Laser therapy	Pain, global health, function, pain pressure threshold, sick leave	PEDro	None	Laser therapy administered with optimal doses can provide short-term pain relief and improve disability	Yes
<b>Boudreault et al. (2014)<sup>16</sup></b>	J Rehabil Med (1.907)	Shoulder	12	Oral NSAIDs	None defined	Cochrane	None	Oral NSAIDs effective at reducing short-term pain but not function	Yes
<b>Buchbinder et al. (2006)<sup>17</sup></b>	J Rheumatol (3.187)	Lateral Elbow	9	ESWT	Pain, function (not pre-defined)	Tool devised by authors	Cochrane Musculoskeletal Group	ESWT provides little to no benefit in pain and function	Yes
<b>Catapano et al. (2020)<sup>18</sup></b>	PMR (1.902)	Shoulder	5	Dextrose Prolotherapy	None defined	Cochrane	None	Prolotherapy is potentially useful adjunct to physical therapy	No
<b>Challoumas et al. (2019a)<sup>19</sup></b>	BMJ Open Sport Exerc Med (1.510)	All	12	Surgery	Pain, function, ROM, treatment	Cochrane	Cochrane Back Review Group	Surgery superior to no treatment/placebo but not	No

					success			sham surgery or physiotherapy	
<b>Challoumas et al. (2019b)<sup>20</sup></b>	Br J Sports Med (11.645)	All	10	Topical GTN	Pain, tenderness, function, strength, ROM, patient satisfaction	Cochrane	Cochrane Back Review Group	Topical GTN superior to placebo in medium term	No
<b>Chen et al. (2019)<sup>21</sup></b>	Arthroscopy (4.433)	Patellar	11	Non-surgical treatments	Pain, function	PEDro	None	LR-PRP is most effective non-surgical treatment	Yes
<b>Coombes et al. (2010)<sup>22</sup></b>	Lancet (59.102)	All	41	Corticosteroid and other injections	Pain, function, improvement	PEDro	Cochrane Back Review Group	Corticosteroid injections are effective in the short-term, other injections may provide long-term benefit for lateral elbow tendinopathy	Yes
<b>Dan et al. (2019)<sup>23</sup></b>	Cochrane Database Syst Rev (7.755)	Patellar	2	Surgery	Pain, function, satisfaction, QoL, withdrawals, adverse events,	Cochrane	GRADE	Inconclusive due to low quality of evidence; surgery likely no more effective than eccentric	Yes

					tendon rupture, return to sport			exercise	
<b>de Vos et al. (2014)<sup>24</sup></b>	Br J Sports Med (11.645)	Lateral Elbow	6	PRP	Pain, function	PEDro	Cochrane Back Review Group	PRP not effective	No
<b>Desjardins-Charbonneau et al. (2015a)<sup>25</sup></b>	Int J Sports Phys Ther (2.55)	Shoulder	10	Taping	Pain, function, ROM, strength	Cochrane	None	Inconclusive due to low quality of evidence	Yes
<b>Desjardins-Charbonneau et al. (2015b)<sup>26</sup></b>	J Orthop Sports Phys Ther (3.058)	Shoulder	21	Manual therapy	Pain, function	Cochrane	None	Manual therapy may decrease pain but it is unclear if it improves function	Yes
<b>Desmeules et al. (2016a)<sup>27</sup></b>	J Occup Health (1.800)	Shoulder	10	Exercise	Work-related outcomes	Cochrane	None	Exercise is effective at treating workers and promotes return to work	No
<b>Desmeules et al. (2016b)<sup>28</sup></b>	Physiotherapy (2.534)	Shoulder	6	TENS	Pain, ROM (not pre-defined)	Cochrane	None	Inconclusive due to low quality of evidence	No
<b>Desmeules et al. (2015)<sup>29</sup></b>	Phys Ther Sport (2.000)	Shoulder	11	Therapeutic US	Pain, function	Cochrane	None	Therapeutic US administered with exercise no	Yes

								more superior than exercise alone. Compared to laser treatment it is less effective at alleviating pain	
<b>Dong et al. (2016)<sup>30</sup></b>	Br J Sports Med (11.645)	Lateral Elbow	27	Injection therapies	Pain	Cochrane	None	Some injection therapies can be effective (e.g. BOTOX and PRP) but not corticosteroids. Hyaluronate and prolotherapy need more research.	Yes
<b>Dupley &amp; Charalambous (2017)<sup>31</sup></b>	Knee Surg Relat Res (No IF)	Patellar (refractory)	2	PRP	Function	None	None	PRP superior to dry needling and ESWT	Yes
<b>Fitzpatrick et al. (2017)<sup>32</sup></b>	Am J Sports Med (6.093)	All	18	PRP	Pain	Cochrane	None	Good evidence to support single injection of PRP under US guidance	Yes
<b>Haslerud et al. (2015)<sup>33</sup></b>	Physiother Res Int (1.33)	Shoulder	17	Laser therapy	Pain, global improvement, function	PEDro	None	Laser therapy can offer clinically	Yes



								relevant pain relief and improvement in symptoms alone and in combination with physiotherapy	
<b>Heales et al. (2020)<sup>34</sup></b>	Musculoskel et Sci Pract (2.319)	Lateral Elbow	7	Forearm and wrist orthoses	Pain, function	Joanna Briggs Institute	GRADE	Orthoses can improve pain and pain-free grip strength but not maximal grip strength	No
<b>Huisstede et al. (2011)<sup>35</sup></b>	Man Ther (2.158)	Shoulder (calcific and non-calcific)	17	ESWT	Pain, function, recovery	Furlan et al. (2009)	Cochrane Back Review Group	High-energy ESWT effective for calcific but not non-calcific tendinopathy	No
<b>Ioppolo et al. (2013)<sup>36</sup></b>	Arch Phys Med Rehabil (2.697)	Shoulder calcific	6	ESWT	Clinical improvement (function), resorption of deposits	PEDro	None	ESWT effective in terms of pain, function and resorption of calcific deposits	Yes
<b>Krey et al. (2015)<sup>37</sup></b>	Phys Sportsmed (1.874)	All	4	Tendon needling	None defined	None	None	Tendon needling and possible PRP improve patient-	No

								reported outcome measures	
<b>Lafrance et al. (2019)<sup>38</sup></b>	BMJ Open Sport Exerc Med (1.510)	Shoulder calcific	3	US-guided lavage	Pain, function, calcification size	Cochrane	None	Ultrasound-guided lavage is more effective than shockwave therapy or a corticosteroid injection alone	No
<b>Larsson et al. (2012)<sup>39</sup></b>	Knee Surg Sports Traumatol Arthrosc (3.149)	Patellar	13	All	None defined or used	Cochrane Back Review Group	Cochrane Back Review Group	Eccentric training most effective, US therapy not effective. Surgery, sclerosing injections and ESWT should be further investigated	No
<b>Lee et al. (2011)<sup>40</sup></b>	J Should Elb Surg (2.865)	Shoulder calcific	9	ESWT	Pain, function	PEDro	NHMRC	Inconclusive due to low quality of evidence	No
<b>Li et al. (2019)<sup>41</sup></b>	Medicine (Baltimore) (1.870)	Lateral Elbow	7	PRP, corticosteroid injection	Pain, function	Cochrane	None	Corticosteroid injection superior to PRP in short-term	Yes

								but PRP more effective in long-term	
<b>Liao et al. (2018)<sup>42</sup></b>	Am J Phys Med Rehabil (1.908)	Lower limb tendinopathies	29	ESWT	Pain, function	PEDro	None	ESWT is effective for pain and function	Yes
<b>Lin et al. (2020)<sup>43</sup></b>	Diagnostics (Basel) (2.489)	Shoulder	5	PRP	Pain, function	Cochrane	None	PRP may be beneficial for long-term pain	Yes
<b>Lin et al. (2019)<sup>44</sup></b>	Arch Phys Med Rehabil (2.697)	Shoulder	7	Injection therapies	Pain, function	Cochrane	None	Corticosteroid effective in short but not long-term, PRP and prolotherapy superior in the long-term	Yes
<b>Lin et al. (2018)<sup>45</sup></b>	Clin Rehabil (2.738)	Lateral Elbow	6	Botulinum toxin injection (BOTOX)	Pain, function	Cochrane	None	BOTOX injections superior to placebo and as effective as corticosteroid injections (though less effective for short-term pain)	Yes
<b>Littlewood et</b>	Physiothera	Shoulder	4	Exercise	Pain,	Cochrane	Cochrane	Evidence	No

<b>al. (2012)<sup>46</sup></b>	py (2.534)				disability	Back Review Group	Back Review Group	supports use of exercise but study largely inconclusive	
<b>Louwerens et al. (2014)<sup>47*</sup></b>	J Should Elb Surg (2.865)	Shoulder calcific	20	Minimally invasive therapies	Pain, function, resorption rate	Cochrane	GRADE	High-energy ESWT safe and effective in short- and mid- term	Yes
<b>Maffulli et al. (2015)<sup>48</sup></b>	Br Med Bull (2.804)	Achilles	13	Pharmacologi cal interventions	“Clinical or functional outcomes”	Modified Coleman	None	Inconclusive due to low quality of evidence	No
<b>Magnussen et al. (2009)<sup>49</sup></b>	Clin J Sport Med (2.702)	Achilles	16	Non- operative treatment	Pain, tenderness, tendon thickness, function	None	None	Eccentric exercise has the most evidence for effectiveness. ESWT, injections of corticosteroid, sclerosing agents or deproteinised haemodialysate and topical GTN need more research	No
<b>Martimbianco et al. (2020)<sup>50</sup></b>	Clin Rehabil (2.738)	Achilles	4	Laser therapy	Function, pain, QoL,	Cochrane	GRADE	Inconclusive due to low	Yes

					adverse events			quality of evidence	
<b>Mendonca et al. (2020)</b> <sup>51</sup>	Br J Sports Med (11.645)	Patellar	9	Conservative treatment	Pain, function	PEDro	GRADE	Inconclusive due to low quality of evidence	Yes
<b>Miller et al. (2017)</b> <sup>52</sup>	BMJ Open Sport Exerc Med (1.510)	All	16	PRP	Pain (not pre-defined)	Cochrane	None	PRP more efficacious than control	Yes
<b>Mohamadi et al. (2017)</b> <sup>53</sup>	Clin Orthop Relat Res (4.154)	Shoulder	14	Corticosteroid injections	Pain	Cochrane	None	Corticosteroid injections provide minimal transient pain relief in a small number of patients	Yes
<b>Murphy et al. (2019)</b> <sup>54</sup>	Br J Sports Med (11.645)	Achilles	7	Heavy eccentric calf training (HECT)	Pain, function	Cochrane RoB 2	GRADE	HECT may be superior to no treatment and traditional physiotherapy but inferior to other exercise interventions	Yes
<b>Nogueira &amp; Moura (2015)</b> <sup>55</sup>	Acta Ortop Bras (0.570)	All	3	Laser therapy	Pain (not pre-defined)	None	None	Laser therapy improves pain in Achilles tendinopathy	No
<b>Ortega-Castillo</b>	J Sci Med	Shoulder &	12	Eccentric	Pain,	PEDro	Cochrane	Eccentric	No

<b>&amp; Medina-Porqueres (2016)<sup>56</sup></b>	Sport (3.623)	Lateral elbow		exercise	strength		Back Review Group	exercise effective for pain and strength but its effectiveness compared to other treatments remains questionable	
<b>Sussmilch-Leitch et al. (2012)<sup>57</sup></b>	J Foot Ankle Res (1.604)	Achilles	19	Physical therapies	Pain, function	Modified PEDro, Cochrane	None	Eccentric exercise recommended as first line with or without laser therapy. ESWT may be equally effective	Yes
<b>Tsikopoulos et al. (2016)<sup>58</sup></b>	Phys Ther Sport (2.000)	All	5	PRP	Pain, functional disability	Cochrane	None	PRP provided no more clinical benefit than placebo or dry needling	Yes
<b>Toliopoulos et al. (2014)<sup>59</sup></b>	Clin Rheumatol (2.293)	Shoulder	15	Surgery	Pain, function, ROM (not pre-defined)	Cochrane	None	Surgery no more effective than exercises. Arthroscopic surgery may be superior to open for some	No



								outcome measures	
<b>Vander Doelen &amp; Jelley (2020)<sup>60</sup></b>	J Sci Med Sport (3.623)	Patellar	9	Non-surgical treatments	Pain, function	None	None	Isometric exercise, patellar strapping, sports taping, eccentric exercise, injections with PRP, ABI, and saline and DN demonstrated a short-term pain relieving and functional improvement	No
<b>Verstraelen et al. (2014)<sup>61</sup></b>	Clin Orthop Relat Res (4.154)	Shoulder calcific	5	ESWT	Function, radiological outcomes	Furlan et al. (2009)	None	High-energy more effective than low-energy ESWT for function and deposit resorption	Yes
<b>Wasielewski &amp; Kotsko (2007)<sup>62</sup></b>	J Athl Train (2.253)	Lower Limb tendinopathies	11	Eccentric exercise	Pain, strength	PEDro	None	Eccentric exercise may improve pain and strength	No
<b>Woodley et al. (2007)<sup>63</sup></b>	Br J Sports Med	All	11	Eccentric exercise	Pain, function,	PEDro, Cochrane	Cochrane Back Review	Inconclusive due to low	No

	(11.645)				satisfaction, return to work	Back Review Group	Group	quality of evidence	
<b>Wu et al. (2017)<sup>64</sup></b>	Arch Phys Med Rehabil (2.697)	Shoulder calcific	14	Non-operative treatments	Pain, function, calcification clearance	Cochrane, PEDro	None	US-guided needling and ESWT (radial and high-energy focused) alleviate pain and achieve complete resolution of calcium deposits	Yes
<b>Xiong et al. (2019)<sup>65</sup></b>	Phys Sportsmed (1.874)	Lateral Elbow	4	ESWT vs Corticosteroid	Pain, function	Jadad, Cochrane	None	ESWT may be superior to corticosteroids	Yes
<b>Yan et al. (2019)<sup>66</sup></b>	J Orthop Surg Res (1.907)	Lateral Elbow	5	US therapy and ESWT	Pain, function, pain-free grip strength	Modified Jadad, Cochrane	None	ESWT superior to US therapy up to 6 months for pain and pain-free grip strength	Yes
<b>Zhang et al. (2019)<sup>67</sup></b>	Medicine (Baltimore) (1.870)	Shoulder calcific	8	US-guided lavage	Pain, function, size of calcium deposition, calcification clearance	Cochrane	None	US-guided lavage may be superior to ESWT in pain relief and calcification clearance	Yes

**Suppl. Table 1: Characteristics of included systematic reviews and their key findings.** ESWT, extracorporeal shock-wave therapy; GTN, glyceryl trinitrate; PRP, platelet-rich plasma; ROM, range of movement; TENS, transcutaneous electrical nerve stimulation; QoL, quality of life; US, ultrasound  
\*one non-randomised trial included

**Supplementary Figure 2: Useful statistical formulas**

**\*Difference in means (SD) = Mean<sub>2</sub> – Mean<sub>1</sub> ( $\frac{SD_1 + SD_2}{2}$ )**

Mean<sub>1</sub> (SD<sub>1</sub>) = Mean (SD) at baseline

Mean<sub>2</sub> (SD<sub>2</sub>) = Mean (SD) at follow up

\* Not the same as mean difference (MD), which is the mean value of the combined differences in means

$$SD_{\text{pooled}} = \sqrt{\frac{[SD_1^2(n_1-1)] + [SD_2^2(n_2-1)] + \dots + [SD_k^2(n_k-1)]}{(n_1 + n_2 + \dots + n_k - k)}}$$

SD<sub>1</sub> = SD from study 1, SD<sub>2</sub> = SD from study 2 etc

n<sub>1</sub> = sample size from study 1, n<sub>2</sub> = sample size from study 2 etc

k=number of samples/studies

**Optimal Information Size (OIS)**

$$N = \frac{2(a + b)^2 SD^2}{(x_1 - x_2)}$$

N=the sample size required in each of the groups

x<sub>1</sub> – x<sub>2</sub> = minimal clinically relevant difference (defined by authors)

SD<sup>2</sup> = population variance (SD<sub>pooled</sub> can be used)

a = 1.96 (for 5% type I error)

b = 0.842 (for 80% power)

Required statistic	Available statistic	Conversion
SD	SEM	$SD = SEM \times \sqrt{n}$
	IQR	$SD = IQR / 1.35$
	Range	$SD = \text{length of range} / 4$
	CI	$(\text{length of CI} / 3.92) \times \sqrt{n}$
95% CI	Mean <sub>1</sub> (SD <sub>1</sub> ) and mean <sub>2</sub> (SD <sub>2</sub> )	$CI = \text{mean}_2 - \text{mean}_1 \pm 1.96(SD_{\text{pooled}} / n)$
Mean	Median	Mean = Median

**Useful statistical conversions. CI, confidence interval; IQR, interquartile range; n, sample size; SD, standard deviation; SEM, standard error of mean.**