What are small, medium and large effect sizes for exercise treatments of tendinopathy? A systematic review and meta-analysis

Paul A Swinton, Joanna S C Shim, Anastasia Vladimirovna Pavlova, Rachel Moss, Colin Maclean, David Brandle, Laura Mitchell, Leon Greig, Eva Parkinson, Victoria Tzortziou Brown, Dylan Morrissey, Lyndsay Alexander, Kay Cooper

ABSTRACT

Objective To quantify and describe effect size distributions from exercise therapies across a range of tendinopathies and outcome domains to inform future research and clinical practice through conducting a systematic review with meta-analysis.

Design Systematic review with meta-analysis exploring moderating effects and context-specific small, medium and large thresholds.

Eligibility criteria Randomised and quasi-randomised controlled trials involving any persons with a diagnosis of rotator cuff, lateral elbow, patellar, Achilles or gluteal tendinopathy of any severity or duration.

Methods Common databases, six trial registries and six grey literature databases were searched on 18 January 2021 (PROSPERO: CRD42020168187). Standardised mean difference (SMD_{pre}) effect sizes were used with Bayesian hierarchical meta-analysis models to calculate the 0.25 (small), 0.5 (medium) and 0.75 quantiles (large) and compare pooled means across potential moderators. Risk of bias was assessed with Cochrane’s Risk of Bias tool.

Results Data were obtained from 114 studies comprising 171 treatment arms 4104 participants. SMD_{pre} effect sizes were similar across tendinopathies but varied across outcome domains. Greater threshold values were obtained for self-reported measures of pain (small=0.5, medium=0.9 and large=1.4), disability (small=0.6, medium=1.0 and large=1.5) and function (small=0.6, medium=1.1 and large=1.8) and lower threshold values obtained for quality of life (small=0.2, medium=0.3 and large=0.7) and objective measures of physical function (small=0.2, medium=0.4 and large=0.7). Potential moderating effects of assessment duration, exercise supervision and symptom duration were also identified, with greater pooled mean effect sizes estimated for longer assessment durations, supervised therapies and studies comprising patients with shorter symptom durations.

Conclusion The effect size of exercise on tendinopathy is dependent on the type of outcome measure assessed. Threshold values presented here can be used to guide interpretation and assist with further research better establishing minimal important change.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Exercise therapy, in particular resistance exercise, is frequently used in the management of tendinopathy and is known to have general effectiveness across a range of important outcome domains. There is, however, a lack of research comparing effectiveness across different tendinopathies and outcome domains.

WHAT THIS STUDY ADDS

⇒ This large and comprehensive meta-analysis shows that exercise therapy results in relative-ly wide change distributions relative to baseline. Distributions of standardised mean difference effect sizes appear consistent across the most common tendinopathies. In contrast, substantive differences exist in the distributions of standardised mean difference effect sizes across outcome domains.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ The results of this study provide researchers and clinicians with important information regarding how individuals should be expected to respond to exercise therapy for the management of tendinopathies, thereby influencing decisions regarding the effectiveness of any intervention and how to power future research studies.

BACKGROUND

Tendinopathy is a common musculoskeletal condition associated with degenerative changes and characterised by a combination of pain impaired movement and reduced function that typically requires extended periods for recovery.1-3 Tendinopathy can affect any muscle-tendon unit in the body, however, it is most frequently reported in the Achilles, patellar, lateral elbow, rotator cuff and hip tendons.6 Surveys of prevalence of lower extremity tendinopathy in the general population have reported rates of 11.8 and...
10.5 per 1000 person-years,\textsuperscript{7} while prevalence for upper limb tendinopathies have been estimated between 1.3% and 21.0%\textsuperscript{8-10}. Costs to the individual, the health service and economy (due to absenteeism and loss of productivity) are substantial such that identifying effective interventions is a priority.

Exercise therapy is the mainstay of conservative management of tendinopathy and has focused largely on resistance training, and in many instances eccentric strengthening techniques.\textsuperscript{11} The rationale of exercise therapy is to improve load tolerance and possibly structural adaptation of the musculotendinous unit to restore function.\textsuperscript{12,13} In the early phase of rehabilitation, flexibility exercises are often initiated and incorporated into strengthening regimes to facilitate improvements in mobility.\textsuperscript{11} Effective exercise therapy may also require targeting a range of contributing factors, which not only include muscle weakness and decreased flexibility, but also corticospinal and neuromuscular adaptations resulting from persistent pain.\textsuperscript{14} As such, proprioceptive exercise interventions have been used to retrain normal patterns of muscle recruitment in the rehabilitation of shoulder-related tendinopathies including impingement.\textsuperscript{14-17} The time course of tendinopathy recovery is usually slow, the degree of recovery may be incomplete and there may be differences between tendinopathies and across outcome domains. For example, quality of life may improve less quickly for people with rotator cuff tendinopathy compared with those with Achilles tendinopathy, whereas recovery may be faster for pain and the relative magnitudes of these improvements may not be equivalent. Quantifying any tendinopathy or domain-specific differences in expected improvements would help guide efforts to develop consensus concerning optimal management by enabling better intervention comparisons.

A recent scoping review identified a lack of effective tools to draw general conclusions across the large tendinopathy and exercise therapy research base (~450 primary studies), which featured a wide range of interventions across different tendinopathies, populations and outcome domains.\textsuperscript{18} At present, one of the main tools to synthesise information and therein draw general conclusions include the use of meta-analyses. Most previous meta-analyses have attempted to quantify the effectiveness of interventions using standardised mean difference (SMD) effect sizes and Cohen’s standard benchmarks (small=0.2, medium=0.5 and large=0.8) irrespective of the tendinopathy location, population or outcome domain.\textsuperscript{19-27} Despite Cohen’s recommendations that these general benchmarks should only be used where more relevant context-specific information is unavailable,\textsuperscript{28} use of these standard benchmarks is ubiquitous through behavioural, social and health sciences. However, recent attempts have been made across a range of disciplines to use empirically derived effect size distributions to generate context-specific benchmarks providing better means of establishing the effectiveness of different interventions and drawing general conclusions.\textsuperscript{29-35} Results have frequently demonstrated substantive differences between Cohen’s benchmarks and those derived empirically, with examples of both underestimation and overestimation, and even differences across subdomains within a discipline.\textsuperscript{29} In addition, SMDs are used frequently as a means of informing the minimal important change (MIC) for patients, especially when preferred anchor-based approaches using external criteria such as global ratings of change are not available.\textsuperscript{36} Given the range of tendinopathies and outcome domains commonly investigated, there is potential that the distribution and subsequent appropriate interpretation of therapy effects will be diverse and could benefit from the generation of context-specific benchmarks. Therefore, the purpose of this meta-analysis was to perform a large synthesis of the available research creating empirically derived thresholds to benchmark the effectiveness of exercise therapies and explore potential differences across tendinopathies and outcome domains. The analysis also investigated the potential for moderating effects of commonly reported features including assessment duration of outcomes, therapy supervision (supervised vs unsupervised) and symptom duration of patients. The results of this analysis will provide clinicians and researchers with tendinopathy-specific and domain-specific indicators of effect sizes with which to better interpret intervention outcomes.

METHODS

This meta-analysis is part of a project funded by the National Institute for Health Research; Health Technology Assessment 129388 Exercise therapy for the treatment of tendinopathies. The inclusion criteria were influenced by the project aims, the results of our initial scoping review mapping the exercise and tendinopathy literature, as well as stakeholder workshops. The overall structure of systematic reviews and meta-analyses to address effectiveness were registered in the PROSPERO database (CRD42020168187) and individual full protocols made publicly available prior to any analyses.\textsuperscript{37,38} The review was conducted according to the PRISMA 2020 statement with checklist provided in online supplemental file 1.

Protocol deviations

Multiple protocol deviations occurred due to pragmatic considerations and reflecting on processes from previous work packages in the larger project. Originally, it was intended to extract data not in duplicate but to quantify reliability based on a random 10% sample. Given the large number of reviewers extracting data it was decided to perform extraction in duplicate with agreement following differences when required. Originally, it was intended to conduct risk of bias using the ROBINS-I tool for quasi-experimental studies.\textsuperscript{39} As outlined in the following sections, due to pragmatic considerations Cochrane’s Risk of Bias tool was used for both randomised and non-randomised designs.\textsuperscript{40}
Stakeholder involvement
People with lived experience of exercise for tendinopathy were involved in all stages of the review, from inception to dissemination activities, and were recruited via National Health Service (NHS) public involvement networks and social media. Two people who had received exercise for Achilles tendinopathy and rotator cuff related shoulder pain (RCRSP) in NHS and private settings contributed to the design stage by influencing the review questions. One of these people (female, RCRSP) went on to contribute to the oversight committee throughout the review and assisted with reviewing dissemination materials. We also held a stakeholder workshop to inform the direction of the review; four females with lived experience of RCRSP or patellar tendinopathy, including one high performance athlete, took part. We had anticipated greater public involvement; timing of activities (during COVID-19 lockdown) and conducting them solely online may be contributing factors. Nonetheless, contributions were helpful in informing the review.

Inclusion criteria
Participants
This meta-analysis included people of any age or gender with a diagnosis of RCRSP, lateral elbow, patellar, Achilles or gluteal tendinopathy of any severity or duration. Studies that included participants with tendinopathy in the absence of full thickness or large tears were included. We accepted trial authors’ diagnoses where a clearly verifiable group of clinical features is reported including: pathognomonic location of pain; a symptom altering response to applied load and/or stretch, with there being a specific test for most tendinopathies; strategies to rule out differential diagnoses; ultrasound or MRI confirmation of structural change. We included studies with mixed groups where there was clear reporting of the tendinopathic group, or those participants comprised >90% of the investigated cohort.

Intervention
The intervention being assessed is exercise therapy comprising five different therapy classes: (1) resistance, (2) plyometric, (3) vibration, (4) flexibility and (5) proprioception. Definitions for each therapy class are presented in online supplemental file 2. Interventions combining exercise with other active therapies (eg, laser, shockwave, manual therapy or injection) were not included. We included exercise therapies delivered in a range of settings and delivered by a range of health, exercise professionals or support workers. We also included both supervised and unsupervised exercise therapies. As part of the inclusion criteria, we required studies to report sufficient information regarding the exercise intervention to enable appropriate identification of treatment duration, therapy class and exercise dose. In clinical settings, it has been recommended that exercise dose is determined by duration, frequency and intensity. To be included in the review, we required studies to provide sufficient information to describe at least two of the three exercise dose parameters. Where sufficient information was not presented in the main text of a study, a search was made of the publishers’ website to check for online supplemental files that may include relevant information.

Comparator
No comparators were included, with effect sizes used to quantify the intervention effectiveness of exercise only therapies based on change relative to baseline.

Outcomes
Based on the results of our initial scoping review and subsequent stakeholder workshops,18 we included outcomes that assessed six domains: (1) disability, (2) physical function capacity (PFC), (3) function, (4) pain (on loading/activity, over a specified time or without further specification), (5) quality of life and (6) range of motion (ROM) (shoulder joint only). Definitions of each domain and example tools are presented in online supplemental file 3.

Types of studies
We included randomised controlled trials (RCTs) and non-RCTs where at least one intervention arm comprised an exercise therapy that could be categorised according to the therapy classes outlined.

Context
The context included primary care, secondary care or community locations in nations defined as very high or high on the Human Development Index (top 62 countries at the time of protocol development) for the findings to be relevant to the UK context.41

Exclusion criteria
In addition to coding treatment arm interventions according to exercise criteria, we also coded non-active (eg, wait-and-see, placebo and sham) and non-exercise (eg, electrotherapy, biomechanics, manual therapy, injection therapy and surgery) categories according to preset definitions (online supplemental file 4). Where any of the non-active or non-exercise categories could be assigned to a treatment arm intervention, these were excluded from the review.

Search strategy
The search strategy used for this study was part of a larger search conducted to scope the entire tendinopathy and exercise therapy research base.18 The search comprised three steps. First, a limited search of MEDLINE and CINAHL using initial keywords was conducted to develop a full search strategy. Second, the full search strategy was adapted to each database and applied systematically to: MEDLINE, CINAHL, AMED, EMBase, SPORTDiscus, Cochrane library (Controlled trials, Systematic reviews), JBI Evidence Synthesis, PEDRo and Epistemonikos (search terms for each database are presented in online supplemental files that may include relevant information).
supplemental file 5. In addition, six trial registries, five grey literature data bases and Google Scholar were searched (online supplemental file 5). Finally, the third step involved conducting a search of cited and citing articles using Scopus and handsearching a total of 130 systematic reviews that were identified to include information relevant to exercise therapy and tendinopathy. No limit was placed on language, with research studies reported in languages other than English translated via Google Translate or via international collaborations of the review team members. Searches were initiated from 1998 as (1) the heavy load eccentric calf-training protocol for Achilles tendinosis by Alfredson et al was published in 1998 and may be considered seminal work in the field of tendinopathy and (2) there has been a proliferation of research on exercise interventions for tendinopathies post 1998. The final date of the search was 18 January 2021.

Study selection
Proquest Refworks was used to manage references and remove duplicates before importing to Covidence (Melbourne, Australia) to facilitate screening and initiate a second deduplication process. Titles/abstracts were reviewed, independently, by two members of the research team. Full-text copies of all studies included at title/abstract screening stage were retrieved and also reviewed, independently, by two members of the research team. Conflicts were resolved by discussion or by input from a third reviewer.

Data extraction
Data were extracted independently by eight members of the review team (PAS/KC/LA/EP/LS/S/AVP) into preiloted Excel sheets. Data were independently coded as described in the accompanying extraction codebook (online supplemental file 6) by two members from the review team. Differences in entries were detected through automatic checking in MS Excel and then agreed between the same two reviewers. Where pre-post intervention data were not presented in text but in figures, data were extracted using digitisation software (PlotDigitizer V.2.6.8 Windows).

Risk of bias
We used Cochrane’s Risk of Bias tool and six outcome domains: (1) selection bias (random sequence generation and allocation concealment), (2) performance bias (blinding of participants), (3) detection bias (blinding of outcome assessors), (4) attrition bias (incomplete outcome data), (5) reporting bias (selective reporting) and (6) other bias, to assess risk of bias for RCTs and domains 2–6 for non-randomised trials. Risk of bias was recorded for each outcome and time point within each study. When obtaining a summary risk of bias for each domain within a study, the mode category across all outcomes and timepoints was selected. The Cochrane’s Risk of Bias tool was selected as a recent review of popular risk of bias tools in tendinopathy management highlighted none were superior and Cochrane’s Risk of Bias tool could be semiautomated with RobotReviewer, a machine learning system software. RobotReviewer was used to make initial assessments on selection bias and performance bias domains, with manual validation made on the relevant free texts extracted to support the final selection of low, high or unclear risk of bias. This semiautomated process was more efficient, and pragmatic given the large number of included studies and provided an additional element of consistency in the review process.

Statistical analysis overview
In general, meta-analyses attempt to pool data across studies quantifying effectiveness by focusing on a weighted mean value. Due to the use of different outcome domains and different tests within a specific outcome domain, pooling of data requires initial standardisation. The most common metric used to standardise effectiveness is the SMD effect size, which divides the group change by the preintervention SD. Meta-analyses seek to describe the underlying population effect size based on a normal distribution with the mean representing the most likely value across studies, and the SD representing the dispersion that can be expected across individual studies. However, a focus on a single central value provides limited description of the overall distribution that can be expected. By providing estimates of the 0.25 quantile (the value which 25% of observed results are expected to be below), the 0.5 quantile and the 0.75 quantile, a more detailed description is obtained. In addition, these estimates can be used to provide benchmarks to interpret the effectiveness of future interventions, with the traditional qualitative labels of ‘small’ (0.25 quantile), ‘medium’ (0.5 quantile) and ‘large’ (0.75 quantile) used to provide a simple scale.

Most meta-analyses are conducted within a frequentist framework where a focus is placed on the mean value and hypothesis testing with CIs to identify whether they overlap a zero effect. In contrast, analyses performed within a Bayesian framework are more flexible enabling, for example, quantile values to be estimated and can be interpreted intuitively through reporting of subjective probabilities. Given the purpose of the present meta-analysis to describe the overall distribution of effect sizes and explore differences among potentially relevant factors (eg, tendinopathy location and outcome domains), a Bayesian framework was selected. Finally, many meta-analyses only extract a single outcome from each study per-analysis to avoid complexities due to relationships within data. However, more precise estimates of effects may be obtained by including all relevant data but accounting for covariances of multiple study outcomes within the meta-analysis model. A recommended approach to account for hierarchical structures (eg, multiple measurements from the same study, multiple measurements made from difference outcomes within the same study and multiple measurements made from
the same outcome in the same study) is to apply multi-level meta-analysis models, which were included for the present study.

**Statistical analysis details**

SMD<sub>pre</sub> effect sizes were calculated by dividing the relevant mean difference by the preintervention SD and including a small sample bias correction. Where required, SMD<sub>pre</sub> values were reflected by multiplying by –1 to ensure that positive values represented an improved clinical effect. Where outcomes were assessed at multiple time points following baseline measurement, all possible SMD<sub>pre</sub> values were calculated and included in the meta-analysis models. Where means and SDs were not presented but included combinations of the median, range or IQR, values were estimated by the calculations presented by Wan et al. Where sufficient information was not available to estimate SDs, these were imputed through simple linear regression of the log transformed SD and means obtained from all other studies. Separate regressions were performed for preintervention and postintervention data.

All meta-analyses were conducted using a nested four-level model. The series of nestings included the individual study (level 4), the outcome (level 3), the measurement occasion (level 2) and the sampling variance (level 1). Standard distributional assumptions were used to calculate the sampling variance of SMD<sub>pre</sub> values. However, the calculation requires an estimate of the pre–post correlation which is rarely reported in studies. To account for uncertainty in the sampling variance, values within the model were allowed to vary and were estimated by including an informative Gaussian prior approximating correlation values centred on 0.7 and ranging from 0.5 to 0.9. The relative contributions of variance sources were described by variance partition coefficients (VPCs) which were calculated by dividing each estimated variance level by the total sum. Therefore, the higher the VPC for the outcome and measurement levels the greater the covariances within the data. Analyses were only completed where a minimum of 50 effect sizes were available to appropriately describe distributions. Sensitivity analyses checking the potential influences of study type (RCTs vs quasi-RCTs) and study quality (low risk of bias: >50% ‘low risk’ vs high risk of bias: 50% ‘low risk’) on effect size thresholds were conducted through subset comparisons.

It was determined a priori to assess the influence of tendinopathy location, outcome domain, assessment duration, symptom duration and supervised versus non-supervised exercise on effect sizes. This was achieved by subset comparisons for tendinopathy location and outcome domain, and meta-regressions for assessment duration, symptom duration and supervision. Meta-regressions were presented by selecting one level of the factor as a reference to make comparisons with the median and 95% CrI (β<sub>Reference_Comparison</sub> = Median (95%CrI: lower bound to upper bound), such that β>0 indicates an increased effect of the comparison relative to the reference).

The importance of removing outliers to obtain accurate estimates of meta-analysis parameters was identified in a previous and similar large meta-analysis of exercise SMD<sub>pre</sub> values. Outlier SMD<sub>pre</sub> values were identified by adjusting the empirical distribution by a Tukey g- and-h distribution and obtaining the 0.0125 and 0.9875 quantiles, with values beyond these points removed prior to further analysis. Meta-analyses were conducted using the R wrapper package brms interfaced with Stan to perform sampling. Convergence of parameter estimates were obtained for all models with Gelman-Rubin R-hat values below 1.5. Data used for the analyses and R code are presented in online supplemental files 7 and 8.

**RESULTS**

**Study selection**

The search strategy identified a total of 9246 potential studies, with 4635 remaining following removal of duplicates (figure 1). After title and abstract screening 4210 studies were removed leaving 425 studies obtained for full text screening. Of these studies, a further 311 were excluded based primarily on insufficient description of the exercise stimulus (116 studies) and not including exercise-only treatment arms (75 studies). In total, data from 114 studies (100 RCTs and 14 quasi-experimental) comprising 171 treatment arms and 4104 participants were included in the meta-analyses. A table of includes studies along with reference lists of included and excluded studies are presented in online supplemental files 9 and 10. Risk of bias expressed for each individual study are
presented in online supplemental file 11 with summaries presented in table 1. For RCTs, risk of bias was highest for ‘other bias’ (47% high risk of bias), blinding of participants (42% high risk of bias) and selective reporting (67% unclear risk of bias). For quasi-experimental trials, risk of bias was also highest for ‘other bias’ (77% high risk of bias) and reporting quality was also lower with high percentages of unclear risk of bias identified for selective reporting (85% unclear risk of bias) and blinding (participants: 39% unclear risk of bias; outcome assessors: 61% unclear high risk of bias).

Descriptions of the study characteristics, tendinopathy location and outcome domains are presented in table 2. Outcomes obtained from studies investigating gluteal tendons were not included in the analysis based on the number of effect sizes falling below the a priori threshold set to generate accurate estimates of the population effect size distribution.

**Description of effect size distributions**

From the initial 1454 outcomes extracted, a total of 38 outliers were removed from the analysis with a lower bound threshold of −0.82 (6 effect sizes below) and an upper bound threshold of 7.0 (32 effect sizes above). Across all outcomes and tendinopathy locations, direct calculation of the 0.25 (small), 0.5 (medium) and 0.75 quantiles (large) from the complete empirical data returned the following SMD \( \text{pre} \) values: 0.37, 0.77 and 1.31, respectively. Application of the meta-analysis model across the data with borrowing of information across studies resulted in similar but shrunken estimates (0.25 quantile \( \text{post} \) = 0.34 (95% CrI: 0.31 to 0.37); 0.5 quantile \( \text{post} \) = 0.73 (95% CrI: 0.70 to 0.77) and 0.75-quantile \( \text{post} \) = 1.21 (95% CrI: 1.17 to 1.27)). A forest plot of effect sizes illustrating effect sizes across studies is presented in online supplemental file 12. Sensitivity analyses checking the potential influences of study type and study quality are presented in online supplemental file 13. No evidence was obtained of greater effect sizes with quasi-experimental designs or with studies identified as high risk of bias.

Analyses of effect size distributions across the different tendinopathy locations are illustrated in figure 2 with numerical values presented in table 3. Analyses were
pooled across all outcome domains as a means to compare tendinopathy locations with the largest amount of data possible. Analysis of the modelled small, medium and large thresholds showed considerable overlap in small and medium thresholds across all tendinopathy locations (0.25 quantile \(0.5\) ranged from 0.28 to 0.38; 0.5 quantile \(0.5\) ranged from 0.70 to 0.82). However, greater divergence was identified for large threshold estimates, with the greatest values estimated for the elbow (0.75 quantile \(0.5\) ranged from 1.18 to 1.49).

Analyses of effect size distributions across outcome domains are presented in figure 3 with numerical values presented in table 4. A clear split was identified between the domains of quality of life and the objective measures of PFC and ROM, versus the subjective measures of function, disability and pain. The lowest threshold values were estimated for quality of life, PFC and ROM, with the small threshold for quality of life estimated to be below zero (0.25 quantile \(0.5\) = −0.21 (95% CrI: −0.32 to −0.09)). In contrast, the greatest effect sizes values were obtained for outcomes measuring disability, pain and function with the reduced amount of data for function resulting in wider credible intervals. Central estimates indicated that small threshold estimates for domains with the greatest effect sizes were situated between the medium and large threshold estimates for domains with the lowest effect sizes (figure 3).

### Moderator analyses

Moderator analyses investigating potential changes in the mean effect size across all outcomes and tendinopathies...
are presented in table 5. Evidence of a moderator effect was identified for assessment duration (short: ≤12 weeks, medium: 13–52 weeks and long duration: >52 weeks), with results showing a hierarchy and greater mean estimate with increased time from baseline assessment (β_short:medium0.5 = 0.28 (95% CrI: 0.23 to 0.33), p>0.999; β_short:long0.5 = 0.37 (95% CrI: 0.27 to 0.47), p>0.999). These estimates remained consistent after meta-regressions were controlled for tendinopathy location and outcome domain (β_short:medium0.5 = 0.28 (95% CrI: 0.23 to 0.33), p>0.999; β_short:long0.5 = 0.38 (95% CrI: 0.27 to 0.48), p>0.999). Consistent evidence of a moderating effect was also obtained for exercise supervision, with a greater mean estimate for supervised exercise therapies (β_unsupervised:supervised0.5 = 0.38 (95% CrI: 0.05 to 0.71), p=0.989), which increased in estimate after controlling for tendinopathy location and outcome domain (β_unsupervised:supervised0.5 = 0.66 (95% CrI: 0.35 to 0.98), p>0.999). Finally, some evidence was obtained indicating a hierarchy of effects with regards to symptom duration and greater mean estimates with patients reporting shorter symptom durations (β_medium:short0.5 = 0.33 (95% CrI: −0.12 to 0.81), p=0.920; β_long:short0.5 = 0.51 (95% CrI: 0.08 to 0.96), p=0.989). However, differences between the symptom duration levels were close to zero after controlling for tendinopathy location and outcome domain (β_medium:short0.5 = −0.05 (95% CrI: −0.44 to 0.56), p=0.411; β_long:short0.5 = 0.06 (95% CrI: −0.34 to 0.47), p=0.619).

**DISCUSSION**

The present analysis represents the largest quantitative synthesis of exercise therapy interventions for the management of tendinopathies to date. Data from a total of 114 studies were included, with results demonstrating that substantive improvements in outcomes from baseline are generally obtained. With important clinical relevance, the meta-analyses identified clear differences in the distribution of effect sizes across outcome domains. The greatest values were generally obtained for subjective patient reported outcomes including disability and pain-related outcomes, with considerably lower values obtained for measures of quality of life and objective measures including PFC and ROM. Considerable overlap in effect size distributions were identified across the different tendinopathy location investigated, indicating that similar substantive improvements can be obtained with exercise therapies. Moderator analyses provided consistent evidence of increased improvement in outcomes as time increased from baseline assessment and with supervised compared with non-supervised exercise therapies. Some evidence was obtained for greater

![Figure 3](image-url)

**Figure 3** Effect size distributions across outcome domains with identification of small, medium and large thresholds. Each curve represents density plot of empirical effect size distribution for specific outcome domain. Density curves and effect size thresholds are presented in same colours. PFC, physical function capacity; QoL, quality of life; ROM, range of motion.
improvements with patients reporting symptoms for shorter durations, however, differences did not remain after controlling for outcome domain and tendinopathy location.

Of the 114 studies included in the analysis, 100 (88%) were RCTs and 14 (12%) were quasi-experimental trials. Therapies predominantly focused on resistance exercise, with 70% of outcomes obtained from a treatment arm where this was the dominant class. Tendinopathy has been clinically defined as persistent pain and loss of function related to mechanical loading on a degenerative tendon.53 Therefore, resistance exercise that focuses on restoration of loading ability is established as the mainstay of rehabilitation, particularly for lower limb tendinopathies, and is in keeping with current guidance.54 Across the included studies flexibility and proprioceptive training were frequently combined with resistance exercise. However, these alternative therapy classes were rarely the dominant exercise class, and where they were, this tended to be restricted to management of RCRSP which accounted for over 50% of outcomes measured in the studies. Given the focus on resistance exercise and the desire to change the mechanical properties of the tendon, the standard duration of exercise therapies featured in the included studies is potentially a major limitation. While clear reporting of therapy duration

<table>
<thead>
<tr>
<th>Moderator</th>
<th>Estimate of mean (95% CrI)</th>
<th>Probabilities</th>
<th>Study VPC (75% CrI)</th>
<th>Outcome VPC (75% CrI)</th>
<th>Measurement VPC (75% CrI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assessment duration</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Short (≤12 weeks)</td>
<td>0.97 (0.85 to 1.1)</td>
<td>p(Medium&gt;Short) &gt; 0.999</td>
<td>0.71 (0.66 to 0.75)</td>
<td>0.27 (0.23 to 0.32)</td>
<td>0.02 (0 to 0.04)</td>
</tr>
<tr>
<td>Medium (13–52 weeks)</td>
<td>1.2 (1.1 to 1.4)</td>
<td>p(Long&gt;Medium) = 0.980</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Long (&gt;52 weeks)</td>
<td>1.4 (1.2 to 1.6)</td>
<td>p(Long&gt;Short) &gt; 0.999</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Symptom duration</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 year or less</td>
<td>1.4 (1.1 to 1.7)</td>
<td>p(1yr&gt;2yr) = 0.920</td>
<td>0.81 (0.76 to 0.84)</td>
<td>0.17 (0.14 to 0.21)</td>
<td>0.03 (0.00 to 0.05)</td>
</tr>
<tr>
<td>2 years or less</td>
<td>1.1 (0.74 to 1.4)</td>
<td>p(2yr&gt;4yr) = 0.772</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Over 2 years</td>
<td>0.88 (0.56 to 1.2)</td>
<td>p(1yr&gt;4yr) = 0.989</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supervision</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supervised</td>
<td>1.4 (1.1 to 1.7)</td>
<td>p(Supervised&gt;Unsupervised) = 0.989</td>
<td>0.77 (0.73 to 0.81)</td>
<td>0.20 (0.17 to 0.24)</td>
<td>0.03 (0.00 to 0.05)</td>
</tr>
<tr>
<td>Unsupervised</td>
<td>0.99 (0.83 to 1.2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Crl, credible interval; p, subjective Bayesian probability; VPC, variance partition coefficient.
only occurred in 49% of studies, where it was reported, the median duration was only 8 weeks and over 90% of studies included durations of 12 weeks or shorter. Tendon healing is known to be a complex and lengthy process, with remodelling only beginning 1–2 months postinjury and extending beyond 1 year, suggesting that longer duration exercise interventions are required, which is in keeping with guidelines recommending a minimum 12 weeks duration.55 It is also relevant to note that in clinical practice, education and exercise programmes are frequently provided for patients to continue therapy after the initial intervention which may also contribute to the findings observed here including evidence of greater improvements with increased time from baseline assessment.

Review of previous meta-analyses and individual studies investigating the effectiveness of exercise therapy for tendinopathy shows that Cohen’s standard benchmarks are most frequently used to interpret effectiveness.19−27,56 The results of the present study show that across most outcomes Cohen’s standard benchmarks of 0.2 (small), 0.5 (medium) and 0.8 (large) would tend to result in an overestimation of effectiveness and for quality of life may result in an underestimation.26 Based on the analyses performed in the present meta-analysis, more appropriate thresholds include −0.4 (small), −0.8 (medium) and −1.3 (large), highlighting a systematic shift by one category (eg, what would have been referred to as a medium/large effect is more representative of a small/medium effect). However, the results of the analyses clearly show that effect size distributions are strongly dependent on the outcome domain, with more subjective patient-reported outcomes (eg, pains scales and patient-rated levels of function and disability) producing substantively larger effects compared with more objective assessments (eg, ROM and quantitative measures of PFC). A wide range of outcome measures are used for tendinopathies, with little consensus in the literature to date,58 although this should improve in future with work ongoing to identify core outcome sets for specific tendinopathies. International consensus reported that nine outcomes domains should be considered core for tendinopathy.57 The majority (8/9) of the consensus domains are patient reported, with only PFC being objectively measured in the clinical setting (eg, number of hops/squats, dynamometry). The results of this study highlight that when using patient-centred self-reported outcomes, substantial changes should be expected from most exercise therapies.

The lowest effect sizes were obtained for quality of life outcomes with the small threshold indicating that over 25% of exercise therapies represented by those investigated will result in patients’ reporting a poorer quality of life. However, it is possible that the generic nature of quality of life instruments (eg, EuroQol-5D instrument) insufficiently reflects tendinopathy-specific symptoms or that potential limitations such as ceiling effects limit the usefulness of these instruments. Tendinopathy may be acute but is typically due to chronic overuse and degeneration, exacerbated by overloading; therefore, patients may have developed specific coping strategies over time. Concern was raised when developing core domains that there is no tendon-specific quality of life measure,55 but that the overall well-being of patients was important to assess. Further research is required to better understand the factors that influence quality of life assessments when managing tendinopathies and the best measurement tool to use.

Considerable overlap in the effect size thresholds across the five tendinopathies assessed suggests that exercise therapies commonly used to manage the different tendinopathies result in similar profiles of improvement. This contrasts previous perspectives that responses to interventions are variable across tendons and even within tendon sites.58 It is likely that the interactions between a single exercise therapy, the tendinopathy and the outcome domain are complex and when better understood can improve patient care. The results of the present analysis, however, indicate that exercise therapies commonly used to manage RCRSP, Achilles, patellar and lateral elbow tendinopathies result in overall similar responses with most causing relatively large changes in relation to baseline SDs. A limitation of the analysis was that it pooled data across all outcome domains to provide sufficient information to compare distributions across the tendinopathies. As the amount of primary data increases, more refined analyses should be conducted to better investigate potential differences across the most common tendinopathies and those which are at present are under researched.

Despite almost all exercise therapies being of short duration (≤12 weeks), a substantial proportion of outcomes were measured beyond the intervention at a medium duration (13–52 weeks), and a relatively small number of instances of longer-term follow-up (>52 weeks). Moderator analysis demonstrated an ordered effect with the smallest mean pooled effect size obtained for short durations (1.0 (95% CrI: 0.87 to 1.1)), and evidence of greater pooled means for medium (1.3 (95% CrI: 1.1 to 1.4); p=0.999) and long durations (1.4 (95% CrI: 1.2 to 1.6), p=0.999). While the absolute magnitude of the differences were relatively small (~0.2 to 0.4 increase) and there were only two studies that included long term observations, the finding are in keeping with previous research59 and provide support for longer duration interventions as well as follow-up periods in studies. Moderator analyses also provided evidence of a greater mean pooled effect with supervised compared with unsupervised exercise therapies (βUnsupervised/Supervised0.5 = 0.70 (95% CrI: 0.39 to 1.0), p=0.797). Previous systematic reviews specifically investigating supervised versus unsupervised exercise therapy for the rotator cuff have suggested that both approaches are likely to lead to similar improvements60 61 While previous reviews employed greater control and focused on more homogeneous comparisons and considerably smaller number

Open access
of outcomes, the findings in the present analysis and the consistency of results after controlling for tendinopathy location and outcome, demonstrate that in general supervised exercise therapies are likely to provide clinically meaningful improvements beyond unsupervised exercise therapies. Finally, limited evidence was obtained to indicate that greater improvements may be obtained with patients reporting shorter symptom durations. However, differences in pooled mean estimates were close to zero after controlling for tendinopathy location and outcome domains, indicating the need for further research, and highlighting limitations of large modelling studies such as those included here, where systematic differences in a range of potential moderating factors can bias results and generate spurious associations.

Other limitations to consider when interpreting and evaluating the results from the present study include precision of estimates given the reliance on published data and limitations of effect sizes to identify clinical significance. Whist attempts were made to include results from unpublished studies, none met the inclusion criteria and it is known that biases influencing individual studies such as publication bias can result in misleading results in meta-analyses. Publication bias, however, is most relevant to analyses of pairwise effect sizes, studies employing cohort designs and analyses focusing on null hypothesis testing. When comparing two interventions that are both effective, pairwise comparisons may fail to identify significant differences thereby reducing the chances of dissemination in peer-reviewed journals. In the present study, analyses were conducted on non-controlled effect sizes which are less likely to be influenced by this bias, except in cohort designs that were excluded. Multiple statistical techniques have been developed to correct for potential biases that can skew the results of meta-analyses. In general, these approaches are best implemented in traditional meta-analyses where the mean effect and confidence intervals are used to make statements regarding a null hypothesis. In contrast, in the present study, inferences focused on describing the majority of the effect size distribution and mitigating biases including potential overestimation of effect sizes through a meta-analysis model that shrunk estimates based on borrowing of strength across studies and accounted for dependencies in the data such that large single values from individual studies had less influence. The use of quantiles to describe distributions and attempts to remove outliers is also likely to have reduced the influence of biases. The importance of removing outliers was highlighted in a previous analysis of SMD analyses. Where anchor-based approaches are best implemented in traditional meta-analyses, studies employing cohort designs and analyses focusing on null hypothesis testing. When comparing two interventions that are both effective, pairwise comparisons may fail to identify significant differences thereby reducing the chances of dissemination in peer-reviewed journals. In the present study, analyses were conducted on non-controlled effect sizes which are less likely to be influenced by this bias, except in cohort designs that were excluded. Multiple statistical techniques have been developed to correct for potential biases that can skew the results of meta-analyses. In general, these approaches are best implemented in traditional meta-analyses where the mean effect and confidence intervals are used to make statements regarding a null hypothesis. In contrast, in the present study, inferences focused on describing the majority of the effect size distribution and mitigating biases including potential overestimation of effect sizes through a meta-analysis model that shrunk estimates based on borrowing of strength across studies and accounted for dependencies in the data such that large single values from individual studies had less influence. The use of quantiles to describe distributions and attempts to remove outliers is also likely to have reduced the influence of biases. The importance of removing outliers was highlighted in a previous analysis of SMD analyses. Where anchor-based approaches are best implemented in traditional meta-analyses, studies employing cohort designs and analyses focusing on null hypothesis testing. When comparing two interventions that are both effective, pairwise comparisons may fail to identify significant differences thereby reducing the chances of dissemination in peer-reviewed journals. In the present study, analyses were conducted on non-controlled effect sizes which are less likely to be influenced by this bias, except in cohort designs that were excluded. Multiple statistical techniques have been developed to correct for potential biases that can skew the results of meta-analyses. In general, these approaches are best implemented in traditional meta-analyses where the mean effect and confidence intervals are used to make statements regarding a null hypothesis. In contrast, in the present study, inferences focused on describing the majority of the effect size distribution and mitigating biases including potential overestimation of effect sizes through a meta-analysis model that shrunk estimates based on borrowing of strength across studies and accounted for dependencies in the data such that large single values from individual studies had less influence. The use of quantiles to describe distributions and attempts to remove outliers is also likely to have reduced the influence of biases. The importance of removing outliers was highlighted in a previous analysis of SMD analyses. Where anchor-based approaches are best implemented in traditional meta-analyses, studies employing cohort designs and analyses focusing on null hypothesis testing. When comparing two interventions that are both effective, pairwise comparisons may fail to identify significant differences thereby reducing the chances of dissemination in peer-reviewed journals. In the present study, analyses were conducted on non-controlled effect sizes which are less likely to be influenced by this bias, except in cohort designs that were excluded. Multiple statistical techniques have been developed to correct for potential biases that can skew the results of meta-analyses. In general, these approaches are best implemented in traditional meta-analyses where the mean effect and confidence intervals are used to make statements regarding a null hypothesis. In contrast, in the present study, inferences focused on describing the majority of the effect size distribution and mitigating biases including potential overestimation of effect sizes through a meta-analysis model that shrunk estimates based on borrowing of strength across studies and accounted for dependencies in the data such that large single values from individual studies had less influence. The use of quantiles to describe distributions and attempts to remove outliers is also likely to have reduced the influence of biases. The importance of removing outliers was highlighted in a previous analysis of SMD analyses. Where anchor-based approaches are best implemented in traditional meta-analyses, studies employing cohort designs and analyses focusing on null hypothesis testing. When comparing two interventions that are both effective, pairwise comparisons may fail to identify significant differences thereby reducing the chances of dissemination in peer-reviewed journals. In the present study, analyses were conducted on non-controlled effect sizes which are less likely to be influenced by this bias, except in cohort designs that were excluded. Multiple statistical techniques have been developed to correct for potential biases that can skew the results of meta-analyses. In general, these approaches are best implemented in traditional meta-analyses where the mean effect and confidence intervals are used to make statements regarding a null hypothesis. In contrast, in the present study, inferences focused on describing the majority of the effect size distribution and mitigating biases including potential overestimation of effect sizes through a meta-analysis model that shrunk estimates based on borrowing of strength across studies and accounted for dependencies in the data such that large single values from individual studies had less influence. The use of quantiles to describe distributions and attempts to remove outliers is also likely to have reduced the influence of biases. The importance of removing outliers was highlighted in a previous analysis of SMD analyses. Where anchor-based approaches are best implemented in traditional meta-analyses, studies employing cohort designs and analyses focusing on null hypothesis testing. When comparing two interventions that are both effective, pairwise comparisons may fail to identify significant differences thereby reducing the chances of dissemination in peer-reviewed journals. In the present study, analyses were conducted on non-controlled effect sizes which are less likely to be influenced by this bias, except in cohort designs that were excluded. Multiple statistical techniques have been developed to correct for potential biases that can skew the results of meta-analyses. In general, these approaches are best implemented in traditional meta-analyses where the mean effect and confidence intervals are used to make statements regarding a null hypothesis. In contrast, in the present study, inferences focused on describing the majority of the effect size distribution and mitigating biases including potential overestimation of effect sizes through a meta-analysis model that shrunk estimates based on borrowing of strength across studies and accounted for dependencies in the data such that large single values from individual studies had less influence. The use of quantiles to describe distributions and attempts to remove outliers is also likely to have reduced the influence of biases. The importance of removing outliers was highlighted in a previous analysis of SMD analyses. Where anchor-based approaches are best implemented in traditional meta-analyses, studies employing cohort designs and analyses focusing on null hypothesis testing. When comparing two interventions that are both effective, pairwise comparisons may fail to identify significant differences thereby reducing the chances of dissemination in peer-reviewed journals. In the present study, analyses were conducted on non-controlled effect sizes which are less likely to be influenced by this bias, except in cohort designs that were excluded. Multiple statistical techniques have been developed to correct for potential biases that can skew the results of meta-analyses. In general, these approaches are best implemented in traditional meta-analyses where the mean effect and confidence intervals are used to make statements regarding a null hypothesis. In contrast, in the present study, inferences focused on describing the majority of the effect size distribution and mitigating biases including potential overestimation of effect sizes through a meta-analysis model that shrunk estimates based on borrowing of strength across studies and accounted for dependencies in the data such that large single values from individual studies had less influence. The use of quantiles to describe distributions and attempts to remove outliers is also likely to have reduced the influence of biases. The importance of removing outliers was highlighted in a previous analysis of SMD analyses. Where anchor-based approaches are best implemented in traditional meta-analyses, studies employing cohort designs and analyses focusing on null hypothesis testing. When comparing two interventions that are both effective, pairwise comparisons may fail to identify significant differences thereby reducing the chances of dissemination in peer-reviewed journals. In the present study, analyses were conducted on non-controlled effect sizes which are less likely to be influenced by this bias, except in cohort designs that were excluded. Multiple statistical techniques have been developed to correct for potential biases that can skew the results of meta-analyses. In general, these approaches are best implemented in traditional meta-analyses where the mean effect and confidence intervals are used to make statements regarding a null hypothesis. In contrast, in the present study, inferences focused on describing the majority of the effect size distribution and mitigating biases including potential overestimation of effect sizes through a meta-analysis model that shrunk estimates based on borrowing of strength across studies and accounted for dependencies in the data such that large single values from individual studies had less influence. The use of quantiles to describe distributions and attempts to remove outliers is also likely to have reduced the influence of biases. The importance of removing outliers was highlighted in a previous analysis of SMD analyses. Where anchor-based approaches are best implemented in traditional meta-analyses, studies employing cohort designs and analyses focusing on null hypothesis testing. When comparing two interventions that are both effective, pairwise comparisons may fail to identify significant differences thereby reducing the chances of dissemination in peer-reviewed journals. In the present study, analyses were conducted on non-controlled effect sizes which are less likely to be influenced by this bias, except in cohort designs that were excluded. Multiple statistical techniques have been developed to correct for potential biases that can skew the results of meta-analyses. In general, these approaches are best implemented in traditional meta-analyses where the mean effect and confidence intervals are used to make statements regarding a null hypothesis. In contrast, in the present study, inferences focused on describing the majority of the effect size distribution and mitigating biases including potential overestimation of effect sizes through a meta-analysis model that shrunk estimates based on borrowing of strength across studies and accounted for dependencies in the data such that large single values from individual studies had less influence. The use of quantiles to describe distributions and attempts to remove outliers is also likely to have reduced the influence of biases. The important
and clinical interpretation of results. All authors contributed to the final manuscript. KC is the overall guarantor of this work.

**Funding** This project is funded by the National Institute for Health Research (NIHR) (Health Technology Assessment (HTA) 129388 Exercise therapy for the treatment of tendinopathies).

**Disclaimer** The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.

**Competing interests** None declared.

**Patient and public involvement** Patients and/or the public were involved in the design, or conduct, or reporting, or dissemination plans of this research. Refer to the Methods section for further details.

**Patient consent for publication** Not applicable.

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

**Data availability statement** All data relevant to the study are included in the article or uploaded as online supplemental information.

**Supplemental material** This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

**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 iDs**
Paul A Swinton http://orcid.org/0000-0001-9663-0696
Joanna S C Shim http://orcid.org/0000-0001-9438-9640
Anastasia Vladimirnova Pavlova http://orcid.org/0000-0002-1633-3411

**REFERENCES**


2. Lewis J. Rotator cuff related shoulder pain, advances in understanding and management. *Journal of Science and Medicine in Sport* 2017;20:47.


29. Schäfer T, Schwarz MA. The meaningfulness of effect sizes in psychological research: differences between sub-disciplines and the impact of potential biases. *Front Psychol* 2019;10:813.


64 Tsujimoto Y, Fuji T, Tsutsumi Y, et al. Minimal important changes in standard deviation units are highly variable and NO universally applicable value can be determined. J Clin Epidemiol 2022;145:92–100.

