HHS Public Access

Author manuscript

J Clin Epidemiol. Author manuscript; available in PMC 2016 November 01.

Published in final edited form as:

J Clin Epidemiol. 2015 November ; 68(11): 1270–1281. doi:10.1016/j.jclinepi.2015.06.006.

Core domain and outcome measurement sets for shoulder pain trials are needed: systematic review of physical therapy trials

Matthew J. Page[a,b], **Joanne E. McKenzie**[a], **Sally E. Green**[a], **Dorcas E. Beaton**[c], **Nitin B. Jain**[d,e], **Mario Lenza**[f], **Arianne P. Verhagen**[g], **Stephen Surace**[h,i], **Jessica Deitch**[h,i], and **Rachelle Buchbinder**[h,i,*]

[a]School of Public Health and Preventive Medicine, Monash University, Level 1, 549 St Kilda Road, Melbourne, Victoria, 3004, Australia [b]School of Social and Community Medicine, University of Bristol, 39 Whatley Road, Bristol, BS8 2PS, UK [c]Department of Occupational Science and Occupational Therapy, University of Toronto, 160-500 University Avenue, Toronto, Ontario, M5G 1V7, Canada [d]Department of Physical Medicine and Rehabilitation, Vanderbilt University School of Medicine, 2201 Children's Way, Suite 1318, Nashville, TN, 37212, USA [e]Department of Orthopaedics, Vanderbilt University School of Medicine, 215 Light Hall, Nashville, TN, 37232, USA [f]Hospital Israelita Albert Einstein, Avenida Albert Einstein, 627/701 - Morumbi, São Paulo, 05652-900, Brazil [g]Department of General Practice, Erasmus MC, University Medical Center Rotterdam, PO Box 2040, Rotterdam, 3000 CA, The Netherlands [h]Monash Department of Clinical Epidemiology, Cabrini Institute, Monash University, 183 Wattletree Road, Malvern, Victoria, 3144, Australia [i]Department of Epidemiology and Preventive Medicine, School of Public Health and Preventive Medicine, Monash University, Level 6, 99 Commercial Road, Melbourne, Victoria, 3004, Australia

Abstract

Objectives—To explore the outcome domains and measurement instruments reported in published randomized controlled trials of physical therapy interventions for shoulder pain (rotator cuff disease, adhesive capsulitis, or nonspecific shoulder pain).

Study Design and Setting—We included trials comparing physical therapy to any other intervention for shoulder pain, indexed up to March 2015 in CENTRAL, MEDLINE, EMBASE, or CINAHL Plus. Two authors independently selected trials for inclusion and extracted information on the domains and measurement instruments reported.

[*]Corresponding author: Monash Department of Clinical Epidemiology, Suite 41, Cabrini Medical Centre, 183 Wattletree Road, Malvern, Victoria, 3144, Australia. Tel.: +61 3 9509 4445; fax: +61 3 9508 1653. rachelle.buchbinder@monash.edu (R. Buchbinder).

Declaration of authorship and conflict of interest: All authors declare to meet the conditions for authorship. M.J.P. and R.B. conceived the study design. J.E.M., D.E.B., S.E.G., N.B.J., M.L., and A.P.V. provided input on the study design. M.J.P., S.S., and J.D. extracted data. M.J.P and R.B. classified outcome measurement instruments into domains. M.J.P. and J.E.M. undertook the statistical analyses. M.J.P. wrote the first draft of the article. All authors contributed to revisions of the article. All authors approved the final version of the submitted article. S.E.G and R.B are authors of two trials included in this review but were not involved in the eligibility assessment or data extraction of these two trials. All other authors declare no competing interests.

Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.jclinepi.2015.06.006.

Results—We included 171 trials. Most trials measured pain (87%), function (72%), and range of movement (67%), whereas adverse events, global assessment of treatment success, strength, and health-related quality of life were measured in 18–27% of trials, and work disability and referral for surgery were measured in less than 5% of trials. Thirty-five different measurement instruments for pain and 29 for function were noted. Measurement of function increased markedly from 1973 to 2014. In rotator cuff disease trials, there was a more frequent measurement of pain and strength and a less frequent measurement of range of movement compared with adhesive capsulitis trials.

Conclusions—There was wide diversity in the domains and measurement instruments reported. Our results provide the foundation for the development of a core domain and outcome measurement set for use in future shoulder pain trials.

Keywords

Outcome assessment (health care); Shoulder pain; Physical therapy modalities; Clinical trial; Systematic review; Research methodology

1. Introduction

Measurement of benefits and harms in randomized trials of health care interventions allows decision makers (patients, clinicians, policy makers) to make evidence-informed choices about health care. To ensure that trials are relevant to decision makers, trialists are encouraged to measure outcome domains (concepts such as pain or function) that are important to patients [1]. Furthermore, domains should be assessed using valid and reliable outcome measurement instruments. That is, tools developed to quantify a domain, such as a visual analog scale (VAS) for pain, which have been shown to perform well in the given context of use [2]. In addition, trialists are encouraged to measure a standardized set of domains to facilitate comparison of results across trials and synthesis of results in meta-analyses [3–5]. However, the domains assessed in clinical trials for many health conditions are not always of most importance to patients, are often inappropriately measured, and are inconsistent across trials [6–13].

To reduce the variation in outcome measurement in trials, “core domain sets” and “core outcome measurement sets” have been developed for several health conditions [14]. Core domain sets are the minimum set of outcome domains recommended for measurement in all trials of a particular condition and thus provide guidance on what domains to measure [2,3,5]. Core outcome measurement sets are the minimum set of measurement instruments that must be administered to cover a corresponding domain and thus provide guidance on how to measure particular domains [2]. Both types of core sets are often developed via consensus methods (eg, the Delphi technique), with participation from patients, health professionals, and researchers. Endorsement of measurement instruments is also often underpinned by an evaluation of the measurement properties of available instruments [14] (note, the term “core outcome set” is also used to refer to the concept of core domain sets, although is broader in scope, encompassing domains such as “pain” along with measures such as “pain relief” and “pain intensity” [3,5]; we have chosen to adopt the Outcome Measures in Rheumatology (OMERACT) filter 2.0 framework [2] terminology for this article).

Evidence of the impact of core domain and outcome measurement sets on trialists' outcome measurement is accruing. In studies comparing outcome domains reported in trials before and after dissemination of a core domain set, greater consistency in domains was observed after dissemination in rheumatoid arthritis trials [15] and ankylosing spondylitis trials [16]. However, publication of a preliminary core domain set for gout had no appreciable impact on the domains measured in subsequent gout trials [17], although this may be related to low statistical power, as only 12 of the 68 trials examined started recruitment after publication of the core domain set. Furthermore, there was still variation in the choice of measurement instruments in trials after dissemination in all three studies, suggesting that adoption of core outcome measurement sets may be more difficult to achieve.

Core domain and outcome measurement sets have been developed for several other musculoskeletal conditions, including low back pain [18,19], and osteoarthritis [20], but not for shoulder pain. Shoulder pain is common, with a reported prevalence of 7–26% in the general population [21]. Shoulder pain is debilitating, impacting on the performance of tasks essential to daily living (such as dressing, personal hygiene, eating, and work), and often results in substantial utilization of health care resources [22–25]. Prior reports have documented how inconsistent the outcome domains and measurement instruments are across intervention studies for shoulder pain [26,27]. Green et al examined 31 randomized trials of interventions for shoulder pain published from 1954 to 1995. Pain and range of movement were measured in 29 and 27 trials, respectively, function was measured in eight (although no trial used a validated disability index), adverse events were measured in nine, and health-related quality of life was not measured in any trial [26]. Rodgers et al examined 28 randomized trials and three nonrandomized studies of interventions for primary (idiopathic) adhesive capsulitis published from 1989 to 2009 and also found great diversity in the domains and measurement instruments selected [27].

To provide a foundation for the development of core domain and outcome measurement sets for use in future shoulder pain trials, several issues require exploration. There has been no systematic evaluation of the domains and measurement instruments in trials of interventions for the most common type of shoulder pain (rotator cuff disease) published after 1995. Furthermore, it is unclear whether the selected domains and measurement instruments in previous trials vary according to the shoulder pain diagnoses examined. For example, it is possible that range of movement may be measured more frequently in trials of interventions for adhesive capsulitis because restriction of passive movement of the shoulder is considered a defining characteristic of that condition [28,29].

The aim of this systematic review was to explore the outcome domains and measurement instruments reported in published randomized controlled trials of physical therapy interventions for rotator cuff disease, adhesive capsulitis, and nonspecific shoulder pain (the most commonly studied shoulder pain diagnoses in clinical trials [26,30]). This measurement review was stimulated by concurrent work on a series of Cochrane reviews investigating the effects of manual therapy and exercise, and electrotherapy modalities, for adhesive capsulitis [31,32] and rotator cuff disease (in progress). The primary objectives of this measurement review were to investigate:

1. the frequency of outcome domains and measurement instruments reported in the trials and
2. whether reporting of outcome domains varied according to the shoulder pain diagnosis and the year of publication.

What is new?

Key findings

- There was wide diversity in the outcome domains and measurement instruments reported in 171 trials of physical therapy interventions for shoulder pain (rotator cuff disease, adhesive capsulitis, or nonspecific shoulder pain). Most trials measured pain (87%), function (72%), and range of movement (67%), whereas adverse events, global assessment of treatment success, strength, and health-related quality of life were measured in 18–27% of trials, and work disability and referral for surgery were measured in less than 5% of trials.
- Thirty-five different measurement instruments for pain and 29 for function were noted across the trials.
- In rotator cuff disease trials, there was a more frequent measurement of pain and strength and a less frequent measurement of range of movement compared with adhesive capsulitis trials. Measurement of function increased markedly from 1973 to 2014.

What this adds to what was known

- The most recent review of outcome domains and measurement instruments in shoulder pain trials included trials published up to 2009. Our study includes trials published up to 2014 and is the first to explore whether reporting of domains varied according to the shoulder pain diagnosis and the year of publication.

What is the implication and what should change now

- We recommend that a core domain and outcome measurement set for shoulder pain trials be developed to reduce diversity of measurement in this important clinical area.

2. Study design and setting

Our study was registered in the Core Outcome Measures in Effectiveness Trials (COMET) database (http://www.comet-initiative.org/studies/details/600), and a full study protocol is available on request.

2.1. Eligibility criteria

2.1.1. Types of studies—We included randomized controlled trials or controlled clinical trials using a quasi-randomized method of allocation (eg, alternate allocation). Reports were eligible regardless of the language, date of publication, or publication status (ie, published or

unpublished). We excluded systematic reviews of trials because we anticipated that an evaluation of trials would be sufficient to inform the need for a core domain and outcome measurement set, given that the domains reported in systematic reviews are often influenced by what has been measured in the included trials.

2.1.2. Types of participants—Numerous diagnostic labels have been used in the literature to describe disorders of the shoulder [30]. On the basis of a previous review of the diagnostic labels and definitions of the study populations [26], most trials could be broadly categorized as examining "rotator cuff disease" (an umbrella term to classify disorders of the rotator cuff, such as subacromial impingement syndrome, rotator cuff tendinopathy or tendinitis, partial or full rotator cuff tear, calcific tendinitis, and subacromial bursitis) or "adhesive capsulitis" (otherwise known as frozen shoulder). We included trials that enrolled participants with rotator cuff disease or adhesive capsulitis, as defined by the trialists. We also included trials that enrolled participants with nonspecific shoulder pain, defined as shoulder pain not explicitly diagnosed as rotator cuff disease or adhesive capsulitis, although not caused by a history of significant trauma, fracture, systemic inflammatory conditions such as rheumatoid arthritis, osteoarthritis, or hemiplegia causing secondary shoulder pain. We excluded trials that enrolled patients with pain in the shoulder region as part of a complex myofascial neck/shoulder/arm pain condition.

2.1.3. Types of interventions—We defined physical therapy interventions as manual therapy (eg, mobilization, manipulation), supervised or home exercises, and electrotherapy modalities (eg, therapeutic ultrasound, laser therapy). Physical therapy interventions could be delivered in isolation (eg, manual therapy alone) or as a multimodal intervention (eg, manual therapy plus exercise plus therapeutic ultrasound). We included trials comparing any physical therapy intervention to placebo, no treatment, or another active intervention (eg, glucocorticoid injection, surgery, or oral medication). Trials comparing two different physical therapy interventions (eg, supervised vs. home exercise program) were also eligible.

2.1.4. Types of outcome domains—We did not consider outcome domains as part of the eligibility criteria; hence, no trial was excluded on the basis of the domains reported.

2.2. Search strategy

We searched the Cochrane Central Register of Controlled Trials (CENTRAL) (to Issue 3, 2015 in *The Cochrane Library*), MEDLINE (January 1966 to March 2015), EMBASE (January 1980 to March 2015), and Cumulative Index to Nursing and Allied Health Literature (CINAHL) Plus (January 1937 to March 2015). The search strategies are presented in Appendix A/Supplementary file at www.jclinepi.com. To identify ongoing or unpublished trials, we searched the database of the World Health Organization International Clinical Trials Registry Platform (ICTRP, http://www.who.int/ictrp/en/), which includes trial registration data sets made available by 15 data providers around the world. To identify any other potentially relevant trials, we reviewed the reference lists of included trials and systematic reviews identified from the electronic searches.

2.3. Selection of trials

Two review authors independently screened all titles and abstracts against a checklist of the predefined eligibility criteria (Section 2.1). If a title or abstract suggested that the trial was eligible, or if there was insufficient information to make a decision, a full-text version of the article was retrieved and assessed for eligibility independently by both authors. Discrepancies were resolved through discussion or adjudication by a third review author. Any reports published in a language other than English were translated using Google translate. If the quality of translation was poor (ie, outcome domains or measurement instruments were indecipherable), trials were excluded.

2.4. Data extraction and management

Two review authors independently extracted data using a standardized data extraction form developed for this measurement review. The authors resolved any discrepancies through discussion or adjudication by a third author, until consensus was reached. We pilot tested the data extraction form and modified it accordingly before use. The data extracted for this measurement review included

- publication characteristics (year of publication);
- participant characteristics (inclusion and exclusion criteria, country, age, sex, and duration of condition in months);
- intervention characteristics (type of physical therapy and duration of intervention in weeks);
- outcome characteristics (domain specified by trialists; measurement instrument used, time interval incorporated in the instrument (eg, pain “over the last 24 hours” vs. “over the last week”), time points at which the instrument was assessed (eg, at the end of 6-week treatment), and whether a primary outcome domain was specified in the trial report).

If any of the previously described data were not clearly presented in the trial report, we contacted the trialists for clarification by sending a maximum of two e-mails separated by 3 weeks. If trialists did not respond, we recorded the relevant data item(s) as “unclear.” One review author compiled all data into a standardized form created in Microsoft Excel. Results of some trials were reported in multiple journal articles. In these instances, we extracted and linked data from all reports, so as to maintain the trial as the unit of analysis.

2.5. Classification of outcome measurement instruments

One author classified each measurement instrument under one of the following nine predefined outcome domains: pain, function, adverse events, global assessment of treatment success, range of movement, strength, health-related quality of life, work disability, and referral for surgery (eg, arthroscopy, manipulation under anesthesia). These domains were nominated because they had been selected as domains of importance for the 2003 Cochrane review of physical therapy interventions for shoulder pain [33]. We classified instruments according to the domain specified by the trialist. Any instrument that could not be classified under the previously described domains was classified as “other.” Classifications were

checked for appropriateness by one author with expertise in shoulder-specific measurement instruments (R.B.). Articles describing the item content and measurement properties of each measurement instrument identified were retrieved to confirm which domain the instrument was originally designed to address. We noted any instances where the domain the instrument was originally designed to address were inconsistent with the domain specified by the trialists.

2.6. Statistical analysis

We summarized results using frequencies and percentages for binary outcomes, and medians and interquartile ranges (IQRs) for continuous outcomes. The nonparametric regression method lowess (locally weighted scatterplot smoothing) was used to smooth the scatterplots of outcome domain use over time (measured in years) [34]. We calculated risk ratios (and their confidence intervals) as a measure of the association between shoulder pain diagnosis (rotator cuff disease or adhesive capsulitis) and reporting of outcome domains. We specifically chose the risk ratio because it is generally more interpretable than the odds ratio [35]. In cases where the outcome domain had been rarely reported, we used penalized likelihood logistic regression to estimate the odds ratio between shoulder pain diagnosis and reporting of outcome domains [36]. Primary analyses were restricted to trials where the authors had explicitly defined patients as having rotator cuff disease or adhesive capsulitis. We undertook post hoc sensitivity analyses to examine if the associations were modified by the inclusion of trials with nonspecific shoulder pain, where we classified these trials as rotator cuff disease trials because this disease is the most common cause of shoulder pain [37,38]. Analyses were undertaken using the functions lowess, cs, and firthlogit in the statistical package Stata version 12 (College Station, Tx) [39].

3. Results

3.1. Search results

A flow diagram of the identification, screening, and inclusion of trials is presented in Fig. 1. The searches yielded 3488 records. After removal of duplicates, the title and abstract of 3166 unique records were screened. Of these, 331 were sought for full-text screening, and 198 reports describing 171 trials were deemed eligible for inclusion in this measurement review.

3.2. Characteristics of included trials

The trials were published between 1973 and 2014 and were conducted in 35 countries, most frequently in Turkey, UK, USA, India, Australia, and Italy (with ≥10 trials published in each of these countries) (Table 1). Participants were diagnosed as having rotator cuff disease (101 [59%] trials), adhesive capsulitis (51 [30%] trials), or nonspecific shoulder pain (19 [11%] trials). The physical therapy interventions varied. The most commonly reported intervention was an electrotherapy modality delivered as an add-on to another physical therapy intervention (26%), followed by manual therapy delivered as an add-on to another physical therapy intervention (21%). Most trials (77%) had two intervention arms. The median sample size was 49 (IQR 30–74; range 8–221).

3.3. Frequency of outcome domains and measurement instruments

In most trials, the domains shoulder pain (87%), function (72%), and range of movement (67%) were measured (Table 2). Adverse events were measured in only 27% of trials. In fewer trials, there was a measurement of global assessment of treatment success (24%), strength (18%), health-related quality of life (18%), work disability (4%), and referral for surgery (2%). Trialists measured a median of three of the nine domains (IQR 2–4; range 1–6).

Outcome domains were assessed using a wide range of measurement instruments across the trials (Table 2; see frequencies (%) of all measurement instruments in Supplementary Table 1/Appendix B at www.jclinepi.com). We noted 35 different pain measurement instruments, with VAS overall pain and VAS pain on movement the most common. Of 137 trials that included a measure of pain using either a VAS or numerical rating scale, most used a 0–10 scale (67%). Across these 137 trials, there were 23 different descriptors for the maximum score on the scale (eg," worst imaginable pain," "severe pain," "intolerable pain"), although in 38% of these trials the descriptor was not reported (see Supplementary Table 2/Appendix B at www.jclinepi.com). A dichotomous measure of pain/no pain was reported in 4% of trials.

The measurement instruments for other domains were also diverse (Supplementary Table 1/Appendix B at www.jclinepi.com). There were 29 measurement instruments noted for function, with the Constant–Murley score [41] and Shoulder Pain and Disability Index [40] the most common. For global assessment of treatment success, most trials used a patient-rated Likert scale, and for health-related quality of life, most trials used subscales of the Short-Form Health Survey (eg, social functioning, vitality, general health) [47]. Range of shoulder movement was measured using either a goniometer or tape measure, although the directions and type of movement assessed (eg, active flexion, passive abduction) varied. There were six instruments noted for strength measurement, with isometric or isokinetic strength via dynamometer the most common.

In 25% of trials, there was at least one measurement instrument that was classified under the "other domain." These instruments were further subclassified into either presence of calcific deposits (measured using radiology), cost or health care use, physical examination maneuvers for rotator cuff disease (eg, Hawkins–Kennedy impingement test), morphologic features of the supraspinatus muscle (measured using radiography), muscle/tendon sensitivity, posture, proprioception, psychological symptoms (eg, depression), scapular dysfunction, severity of the main complaint, shoulder stability, stiffness, supraspinatus tendon thickness (measured using ultrasound), swelling, and weakness on movement. Each was measured in less than 5% of trials. There were no instances where the domain the instrument was originally designed to address was inconsistent with the domain specified by the trialists.

3.4. Frequency of primary outcome domain specification

At least one primary outcome domain was specified in only 47 (27%) trials (Table 3); there was no prioritization of domains as primary or secondary in the remaining (73%) trials. Of

these 47 trials, the primary domains specified were function (60%), pain (19%), global assessment of treatment success (19%), range of movement (13%), and presence of calcific deposits measured using radiography (2%).

3.5. Timing of outcome assessment

In approximately half the trials, domains were measured at up to 3 weeks (51%), between 3 and 6 weeks (50%) and between 6 weeks and 6 months (49%). In only 12% of trials, domains were measured at later than 6 months. Reporting of the time interval incorporated in pain measurement instruments (eg, “current pain” vs. “average pain in the previous week”) varied and was not reported at all in 61% of 148 trials measuring pain (Table 3).

3.6. Association between shoulder pain diagnosis and frequency of outcome domains

Compared with adhesive capsulitis trials, rotator cuff disease trials more frequently included measures of pain (92/101 [91%] rotator cuff disease vs. 39/51 [76%] adhesive capsulitis; risk ratio (RR) 1.2, 95% confidence interval [CI] 1.01–1.4) and strength (26/101 [26%] rotator cuff disease vs. 1/51 [2%] adhesive capsulitis; RR 13.1, 95% CI 1.8–94.0) and less frequently included measures of range of movement (59/101 [58%] rotator cuff disease vs. 42/51 [82%] adhesive capsulitis; RR 0.7, 95% CI 0.6–0.9). For all other domains, the frequency was higher in the rotator cuff disease trials but the 95% CIs included no difference or a lower frequency as possible estimates of effect (Table 4). There were no important changes to the RRs and 95% CIs in sensitivity analyses including nonspecific shoulder pain trials as rotator cuff disease trials (Supplementary Table 3/Appendix B at www.jclinepi.com).

3.7. Frequency of outcome domains over time

Visual inspection of the smoothed scatterplots indicated that measurement of function, pain, strength, and health-related quality of life increased from 1973 to 2014, with function showing the greatest increase (Fig. 2). In contrast, measurement of range of movement and global assessment of treatment success decreased. Measurement of adverse events, work disability, and referral for surgery were similar over this period.

4. Discussion

There was wide diversity in the outcome domains and measurement instruments reported in 171 trials of physical therapy interventions for shoulder pain. In most trials, pain, function, and range of movement were measured, whereas adverse events, global assessment of treatment success, strength, and health-related quality of life were measured in 18–27% of trials, and work disability and referral for surgery were measured in less than 5% of trials. Thirty-five different measurement instruments for pain and 29 for function were noted across the trials. In rotator cuff disease trials, there was a more frequent measurement of pain and strength and a less frequent measurement of range of movement compared with adhesive capsulitis trials. Measurement of function increased markedly from 1973 to 2014.

The trialists’ choice of outcome domains may have been influenced by several factors, including their perceptions about the defining features of the shoulder pain condition

investigated and the type of intervention delivered. For example, on the basis of the restricted range of movement in patients with adhesive capsulitis, some trialists may have considered increasing range of movement of greater priority than increasing strength for this condition, and therefore prioritized measurement of the former. Furthermore, trialists delivering strengthening exercises may have been more inclined to measure the impact of these exercises on patients' strength than trialists delivering range of movement exercises. Both factors (condition and intervention) have likely contributed to the observed diversity in domains across the trials.

There are several possible reasons for the marked increase in function measurement and concomitant decrease in range of movement measurement from 1973 to 2014. Unlike range of movement, very few measurement instruments for shoulder function were available until the mid-1990s, when the Shoulder Pain and Disability Index [40], UK Shoulder Disability Questionnaire [45], DASH [43], and Simple Shoulder Test [51] were published. Furthermore, the measurement patterns may be associated with increasing emphasis on patient-reported outcome measures (PROMs) in clinical decision making, health policy, and reimbursement decisions [52,53]. Patient-reported function instruments provide greater insight into the activities that people with shoulder pain have difficulty with (eg, putting on a jumper, placing or reaching for objects on a high shelf), compared with observer-rated range of movement assessments, and have become increasingly recognized as important patient-relevant outcomes.

The observed diversity in outcome domains suggests that a core domain set for shoulder pain trials is urgently needed, which we have previously asserted [26,31,54,55]. The OMERACT filter 2.0 framework [2] could be used to develop such a core domain set, and OMERACT have supported a special interest group session on this topic to be held at the OMERACT 2016 meeting. In this framework, literature review and consensus process are used to identify at least one core domain within each of three mandatory health "areas," namely death, life impact, and pathophysiological manifestations, and one strongly recommended area, resource use. The nine predefined domains we examined provide a preliminary set to consider in a Delphi consensus process focusing on the life impact or pathophysiological manifestations areas. Important issues to consider are which of the infrequently measured domains in our review (ie, domains other than pain, function, and range of movement) to include in the final core domain set, whether some domains should be excluded because they are not a priority for all types of shoulder pain (eg, range of movement and strength), or whether different core domain sets need to be developed for particular shoulder pain diagnoses (eg, a set for adhesive capsulitis that includes rage of movement and a set for rotator cuff disease that includes strength, with domains such as pain and function included in both). An important stakeholder group to include in discussions about which domains to endorse is patients, whose views on the relative importance of particular domains for different shoulder pain diagnoses may not necessarily match the views of trialists and clinicians.

A core outcome measurement set for shoulder pain trials would be valuable to facilitate consistent selection of measurement instruments in future trials. According to the OMERACT filter 2.0, measurement instruments selected to cover a domain should be

truthful (valid), discriminative, and feasible, and such information should be obtained by a systematic review of the measurement properties of existing instruments [2]. Research of this nature was undertaken in 2014–2015 for PROMs aimed at “activity limitations” of people with shoulder pain [56], PROMs aimed at shoulder function [57], and PROMs measuring any patient-reported domain for use by people with rotator cuff disease [58]. To inform the development of a core outcome measurement set, synthesis of the findings of existing measurement reviews, systematic reviews of the properties of shoulder-specific instruments that have not yet been reviewed, and evaluation of the methodological quality of studies on measurement properties (eg, using the COSMIN checklist [59]) are needed.

In addition to identifying the need for core domain and outcome measurement sets for shoulder pain, we noted several limitations in measurement and reporting in the trials. A primary outcome domain was reported in only 27% of trials. Without such specification, it is difficult to determine how sample size estimation was performed. Furthermore, when there are multiple outcomes with conflicting results, it is difficult for readers to draw a global interpretation without knowing which outcome is the most critical for decision making [60]. In addition, we noted that many trialists did not report the time interval incorporated in measurement instruments for pain, or the descriptor used to denote the maximum score on the pain scale. Assessment of “current pain” vs. “average pain over the last month” provides different information about patients’ condition, and the latter may be more affected by measurement error because of the need to recall pain over a longer period of time. Furthermore, patients may interpret “severe pain” differently from “the worst pain imaginable.” Thus, we echo the guidance of the CONSORT 2010 Statement [60] in recommending that the content of the measurement instruments should be described to allow interpretation of results.

Another limitation of the included trials was that adverse events were measured in only 27% of trials. This finding may relate to the restricted inclusion of trials evaluating physical therapy interventions. Adverse events may occur more often in trials of pharmacologic or surgical interventions, and trialists of physical therapy interventions may anticipate fewer and less serious adverse events. However, reporting of adverse events, including “zero events” and harms that are rare, allows patients and clinicians to determine the benefit-risk profile of an intervention and is therefore essential to include in all trial reports [61,62].

There are several strengths of our study. We followed methods that were predefined and registered in a publicly accessible database (http://www.comet-initiative.org/studies/details/600). We searched for trials indexed in multiple bibliographic databases using a search strategy developed by an information specialist, and two authors independently screened and extracted data from trials. To select trials for inclusion, we applied broad eligibility criteria for the type of condition and intervention, which may enhance the generalizability of our findings. Furthermore, all classifications of outcome measurement instruments were checked for appropriateness by one author in the review team with expertise in shoulder-specific measurement instruments and revised as necessary, which likely reduced the potential for misclassification.

Our study has some limitations. Frequency estimates are a reflection of what was reported, but this may not reflect practice. Previous comparisons between trial protocols and published trials have found that 13–31% of primary outcomes specified in trial protocols were omitted from the publication [63]. Therefore, we may have underestimated the frequency of outcome domains that were actually measured by trialists. Furthermore, we excluded nine potentially eligible articles in a language other than English that we were unable to translate using Google translate, and we did not have access to professional translators. However, these trials comprised only 3% of articles retrieved for full-text screening, so their inclusion would not affect our results. Finally, we only examined trials of physical therapy interventions so our results may not generalize to trials of other interventions for shoulder pain.

5. Conclusions

In conclusion, there was wide diversity in the outcome domains and measurement instruments reported in 171 trials of physical therapy interventions for shoulder pain. This hampers our ability to compare the results of different trials and synthesize the results in systematic reviews. Our findings indicate that a core domain and outcome measurement set for shoulder pain is urgently needed. We will address this need with the support of OMERACT, who have approved a special interest group on shoulder pain at the OMERACT 2016 meeting.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding: There was no direct funding for this research. M.J.P. is supported by an Australian National Health and Medical Research Council (NHMRC) Early Career Fellowship (1088535). R.B. is supported in part by an Australian NHMRC Senior Principle Research Fellowship. J.E.M is supported by an Australian NHMRC Public Health Fellowship (1072366). N.B.J is supported by funding from National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS) (1K23AR059199).

The authors thank Steve McDonald (information specialist) for developing the search strategy. The authors thank Brodwen McBain for independent screening of studies against the eligibility criteria.

References

1. Methodological standards and patient-centeredness in comparative effectiveness research: the PCORI perspective. JAMA. 2012; 307:1636–40. [PubMed: 22511692]
2. Boers M, Kirwan JR, Wells G, Beaton D, Gossec L, d'Agostino M-A, et al. Developing core outcome measurement sets for clinical trials: OMERACT filter 2. 0. J Clin Epidemiol. 2014; 67:745–53. [PubMed: 24582946]
3. Clarke M. Standardising outcomes for clinical trials and systematic reviews. Trials. 2007; 8:39. [PubMed: 18039365]
4. Tugwell P, Boers M, Brooks P, Simon L, Strand V, Idzerda L. OMERACT: an international initiative to improve outcome measurement in rheumatology. Trials. 2007; 8:38. [PubMed: 18039364]

5. Williamson P, Altman D, Blazeby J, Clarke M, Gargon E. Driving up the quality and relevance of research through the use of agreed core outcomes. J Health Serv Res Policy. 2012; 17:1–2. [PubMed: 22294719]
6. Blencowe NS, Strong S, McNair AG, Brookes ST, Crosby T, Griffin SM, et al. Reporting of short-term clinical outcomes after esophagectomy: a systematic review. Ann Surg. 2012; 255(4):658–66. [PubMed: 22395090]
7. Dapuzzo L, Seitz FE, Dodson WC, Stetter C, Kunselman AR, Legro RS. Incomplete and inconsistent reporting of maternal and fetal outcomes in infertility treatment trials. Fertil Steril. 2011; 95:2527–30. [PubMed: 21435640]
8. Endre ZH, Pickering JW. Outcome definitions in non-dialysis intervention and prevention trials in acute kidney injury (AKI). Nephrol Dial Transplant. 2010; 25:107–18. [PubMed: 19812232]
9. Gummesson C, Atroshi I, Ekdahl C. The quality of reporting and outcome measures in randomized clinical trials related to upper-extremity disorders. J Hand Surg. 2004; 29:727–34.
10. Jerosch-Herold C, Leite J, Song F. A systematic review of outcomes assessed in randomized controlled trials of surgical interventions for carpal tunnel syndrome using the International Classification of Functioning, Disability and Health (ICF) as a reference tool. BMC Musculoskelet Disord. 2006; 7:96. [PubMed: 17147807]
11. Kalyoncu U, Dougados M, Daurès J-P, Gossec L. Reporting of patient-reported outcomes in recent trials in rheumatoid arthritis: a systematic literature review. Ann Rheum Dis. 2009; 68:183–90. [PubMed: 18375533]
12. Simpson RC, Thomas KS, Murphy R. Outcome measures for vulval skin conditions: a systematic review of randomised controlled trials. Br J Dermatol. 2013; 169:494–501. [PubMed: 23600623]
13. Sinha IP, Williamson PR, Smyth RL. Outcomes in clinical trials of inhaled corticosteroids for children with asthma are narrowly focussed on short term disease activity. PLoS One. 2009; 4:e6276. [PubMed: 19609438]
14. Gargon E, Gurung B, Medley N, Altman DG, Blazeby JM, Clarke M, et al. Choosing important health outcomes for comparative effectiveness research: a systematic review. PLoS One. 2014; 9:e99111. [PubMed: 24932522]
15. Kirkham J, Boers M, Tugwell P, Clarke M, Williamson P. Outcome measures in rheumatoid arthritis randomised trials over the last 50 years. Trials. 2013; 14:324. [PubMed: 24103529]
16. Bautista-Molano W, Navarro-Compan V, Landewe RB, Boers M, Kirkham JJ, van der Heijde D. How well are the ASAS/OMERACT core outcome sets for ankylosing spondylitis implemented in randomized clinical trials? A systematic literature review. Clin Rheumatol. 2014; 33:1313–22. [PubMed: 24970597]
17. Araújo F, Cordeiro I, Ramiro S, Falzon L, Branco JC, Buchbinder R. Outcomes assessed in trials of gout and accordance with OMERACT-proposed domains: a systematic literature review. Rheumatology (Oxford). 2015; 54:981–93. [PubMed: 25398382]
18. Chiarotto A, Deyo R, Terwee C, Boers M, Buchbinder R, Corbin T, et al. Core outcome domains for clinical trials in non-specific low back pain. Eur Spine J. 2015; 24:1–16. [PubMed: 25510517]
19. Deyo RA, Battie M, Beurskens AJ, Bombardier C, Croft P, Koes B, et al. Outcome measures for low back pain research. A proposal for standardized use. Spine. 1998; 23:2003–13. [PubMed: 9779535]
20. Bellamy N, Kirwan J, Boers M, Brooks P, Strand V, Tugwell P, et al. Recommendations for a core set of outcome measures for future phase III clinical trials in knee, hip, and hand osteoarthritis: consensus development at OMERACT III. J Rheumatol. 1997; 24:799–802. [PubMed: 9101522]
21. Luime JJ, Koes BW, Hendriksen IJM, Burdof A, Verhagen AP, Miedema HS, et al. Prevalence and incidence of shoulder pain in the general population: a systematic review. Scand J Rheumatol. 2004; 33:73–81. [PubMed: 15163107]
22. van der Heijden GJ. Shoulder disorders: a state-of-the-art review. Baillieres Best Pract Res Clin Rheumatol. 1999; 13:287–309. [PubMed: 10952865]
23. Largacha M, Parsons IM IV, Campbell B, Titelman RM, Smith KL, Matsen F III. Deficits in shoulder function and general health associated with sixteen common shoulder diagnoses: a study of 2674 patients. J Shoulder Elbow Surg. 2006; 15:30–9. [PubMed: 16414466]

24. Virta L, Joranger P, Brox J, Eriksson R. Costs of shoulder pain and resource use in primary health care: a cost-of-illness study in Sweden. BMC Musculoskelet Disord. 2012; 13:17. [PubMed: 22325050]
25. Mroz TM, Carlini AR, Archer KR, Wegener ST, Hoolachan JI, Stiers W, et al. Frequency and cost of claims by injury type from a state workers' compensation fund from 1998 through 2008. Arch Phys Med Rehabil. 2014; 95:1048–54. [PubMed: 24480333]
26. Green S, Buchbinder R, Glazier R, Forbes A. Systematic review of randomised controlled trials of interventions for painful shoulder: selection criteria, outcome assessment, and efficacy. BMJ. 1998; 316:354–60. [PubMed: 9487172]
27. Rodgers S, Brealey S, Jefferson L, McDaid C, Maund E, Hanchard N, et al. Exploring the outcomes in studies of primary frozen shoulder: is there a need for a core outcome set? Qual Life Res. 2014; 23:2495–504. [PubMed: 24817317]
28. Reeves B. The natural history of the frozen shoulder syndrome. Scand J Rheumatol. 1975; 4:193–6. [PubMed: 1198072]
29. Binder AI, Bulgen DY, Hazelman BL, Roberts S. Frozen shoulder: a long-term prospective study. Ann Rheum Dis. 1984; 1984:361–4. [PubMed: 6742896]
30. Schellingerhout JM, Verhagen AP, Thomas S, Koes BW. Lack of uniformity in diagnostic labeling of shoulder pain: time for a different approach. Man Ther. 2008; 13:478–83. [PubMed: 18555732]
31. Page MJ, Green S, Kramer S, Johnston RV, McBain B, Chau M, et al. Manual therapy and exercise for adhesive capsulitis (frozen shoulder). Cochrane Database Syst Rev. 2014; 8:CD011275. [PubMed: 25157702]
32. Page MJ, Green S, Kramer S, Johnston RV, McBain B, Buchbinder R. Electrotherapy modalities for adhesive capsulitis (frozen shoulder). Cochrane Database Syst Rev. 2014; 10:CD011324. [PubMed: 25271097]
33. Green S, Buchbinder R, Hetrick S. Physiotherapy interventions for shoulder pain. Cochrane Database Syst Rev. 2003:CD004258. [PubMed: 12804509]
34. Cleveland WS. Robust locally weighted regression and smoothing scatterplots. J Am Stat Assoc. 1979; 74:829–36.
35. Bland JM, Altman DG. Statistics notes. The odds ratio BMJ. 2000; 320:1468. [PubMed: 10827061]
36. Heinze G, Schemper M. A solution to the problem of separation in logistic regression. Stat Med. 2002; 21:2409–19. [PubMed: 12210625]
37. Linsell L, Dawson J, Zondervan K, Rose P, Randall T, Fitzpatrick R, et al. Prevalence and incidence of adults consulting for shoulder conditions in UK primary care; patterns of diagnosis and referral. Rheumatology. 2006; 45:215–21. [PubMed: 16263781]
38. Tekavec E, Joud E, Rittner R, Mikoczy Z, Nordander C, Petersson IF, et al. Population-based consultation patterns in patients with shoulder pain diagnoses. BMC Musculoskelet Disord. 2012; 13:238. [PubMed: 23190941]
39. StataCorp. Stata statistical software: release 12. College Station, TX: StataCorp LP; 2011.
40. Roach KE, Budiman-Mak E, Songsiridej N, Lertratanakul Y. Development of a shoulder pain and disability index. Arthritis Care Res. 1991; 4:143–9. [PubMed: 11188601]
41. Constant CR, Murley AH. A clinical method of functional assessment of the shoulder. Clin Orthop Relat Res. 1987:160–4. [PubMed: 3791738]
42. van der Heijden GJ, Leffers P, Bouter LM. Shoulder disability questionnaire design and responsiveness of a functional status measure. J Clin Epidemiol. 2000; 53:29–38. [PubMed: 10693901]
43. Hudak PL, Amadio PC, Bombardier C. Development of an upper extremity outcome measure: the DASH (disabilities of the arm, shoulder and hand). Am J Ind Med. 1996; 29:602–8. [PubMed: 8773720]
44. Amstutz HC, Sew Hoy AL, Clarke IC. UCLA anatomic total shoulder arthroplasty. Clin Orthop Relat Res. 1981:7–20. [PubMed: 7226634]
45. Croft P, Pope D, Zonca M, O'Neill T, Silman A. Measurement of shoulder related disability: results of a validation study. Ann Rheum Dis. 1994; 53:525–8. [PubMed: 7944638]

46. Hurst H, Bolton J. Assessing the clinical significance of change scores recorded on subjective outcome measures. J Manipulative Physiol Ther. 2004; 27:26–35. [PubMed: 14739871]
47. Ware JE Jr. SF-36 health survey update. Spine. 2000; 25:3130–9. [PubMed: 11124729]
48. EuroQol Group. EuroQol—a new facility for the measurement of health-related quality of life. Health Policy. 1990; 16:199–208. [PubMed: 10109801]
49. Robertson IT, Cooper CL, Williams J. The validity of the occupational stress indicator. Work Stress. 1990; 4:29–39.
50. Beck AT, Steer RA, Carbin MG. Psychometric properties of the Beck Depression Inventory: twenty-five years of evaluation. Clin Psychol Rev. 1988; 8(1):77–100.
51. Lippitt SB, Harryman DT, Matsen FA. A practical tool for evaluation of function: the Simple Shoulder Test. The shoulder: a balance of mobility and stability. The American Academy of Orthopaedic Surgeons. 1993:501–18.
52. Calvert M, Blazeby J, Altman DG, et al. Reporting of patient-reported outcomes in randomized trials: the CONSORT PRO extension. JAMA. 2013; 309:814–22. [PubMed: 23443445]
53. Macefield R, Jacobs M, Korfage I, Nicklin J, Whistance R, Brookes S, et al. Developing core outcomes sets: methods for identifying and including patient-reported outcomes (PROs). Trials. 2014; 15:49. [PubMed: 24495582]
54. Buchbinder R, Green S, Youd JM, Johnston RV. Oral steroids for adhesive capsulitis. Cochrane Database Syst Rev. 2006; 4:CD006189. [PubMed: 17054278]
55. Buchbinder R, Green S, Youd JM, Johnston RV, Cumpston M. Arthrographic distension for adhesive capsulitis (frozen shoulder). Cochrane Database Syst Rev. 2008; 1:CD007005. [PubMed: 18254123]
56. Thoomes-de Graaf M, Scholten-Peeters GGM, Schellingerhout JM, Bourne AM, Buchbinder R, Koehorst M, et al. Evaluation of measurement properties of patient reported outcome measures aimed at “activity limitations” for patients with shoulder pain: a systematic review. submitted 14 Oct 2014.
57. Schmidt S, Ferrer M, González M, González N, Valderas JM, Alonso J, et al. Evaluation of shoulder-specific patient-reported outcome measures: a systematic and standardized comparison of available evidence. J Shoulder Elbow Surg. 2014; 23:434–44. [PubMed: 24406123]
58. Huang, H.; Grant, JA.; Miller, BS.; Mirza, FM.; Gagnier, JJ. A systematic review of the psychometric properties of patient-reported outcome instruments for use in patients with rotator cuff disease. Am J Sports Med. 2015. http://dx.doi.org/10.1177/0363546514565096
59. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res. 2010; 19:539–49. [PubMed: 20169472]
60. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, et al. CONSORT 2010 Explanation and Elaboration: updated guidelines for reporting parallel group randomised trials. BMJ. 2010; 340:c869. [PubMed: 20332511]
61. Ioannidis JP, Evans SJ, Gotzsche PC, O’Neill RT, Altman DG, Schulz K, et al. Better reporting of harms in randomized trials: an extension of the CONSORT statement. Ann Intern Med. 2004; 141:781–8. [PubMed: 15545678]
62. Saini P, Loke YK, Gamble C, Altman DG, Williamson PR, Kirkham JJ. Selective reporting bias of harm outcomes within studies: findings from a cohort of systematic reviews. BMJ. 2014; 349:g6501. [PubMed: 25416499]
63. Dwan K, Gamble C, Williamson PR, Kirkham JJ. the Reporting Bias Group. Systematic review of the empirical evidence of study publication bias and outcome reporting bias—an updated review. PLoS One. 2013; 8:e66844. [PubMed: 23861749]

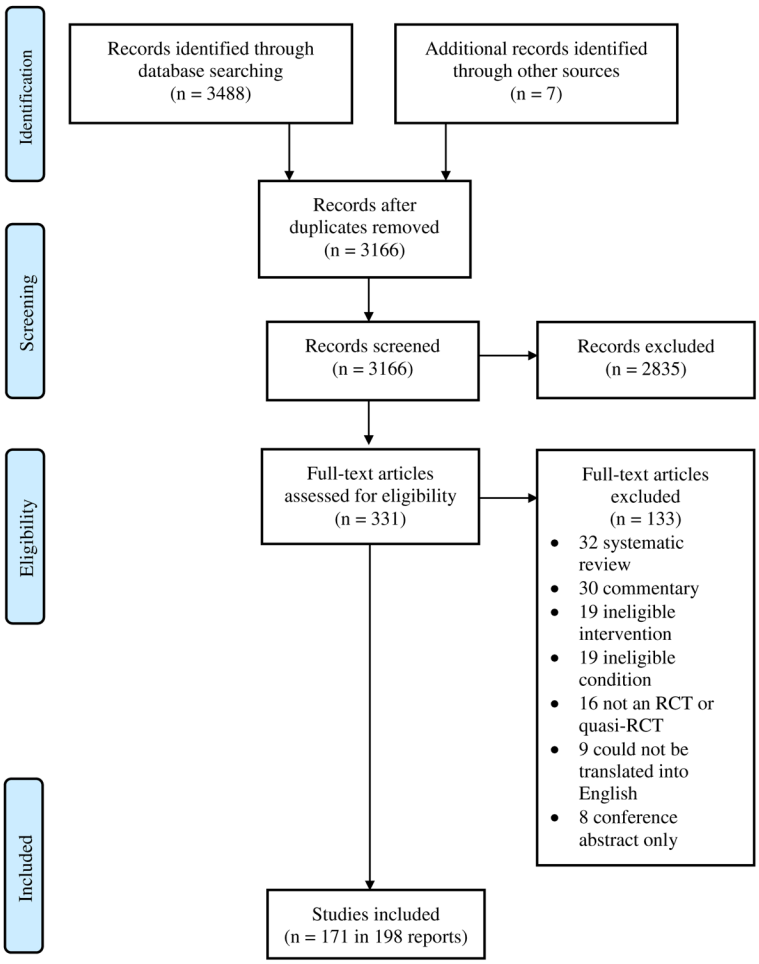

Fig. 1.
Flow diagram of identification, screening, and inclusion of trials.

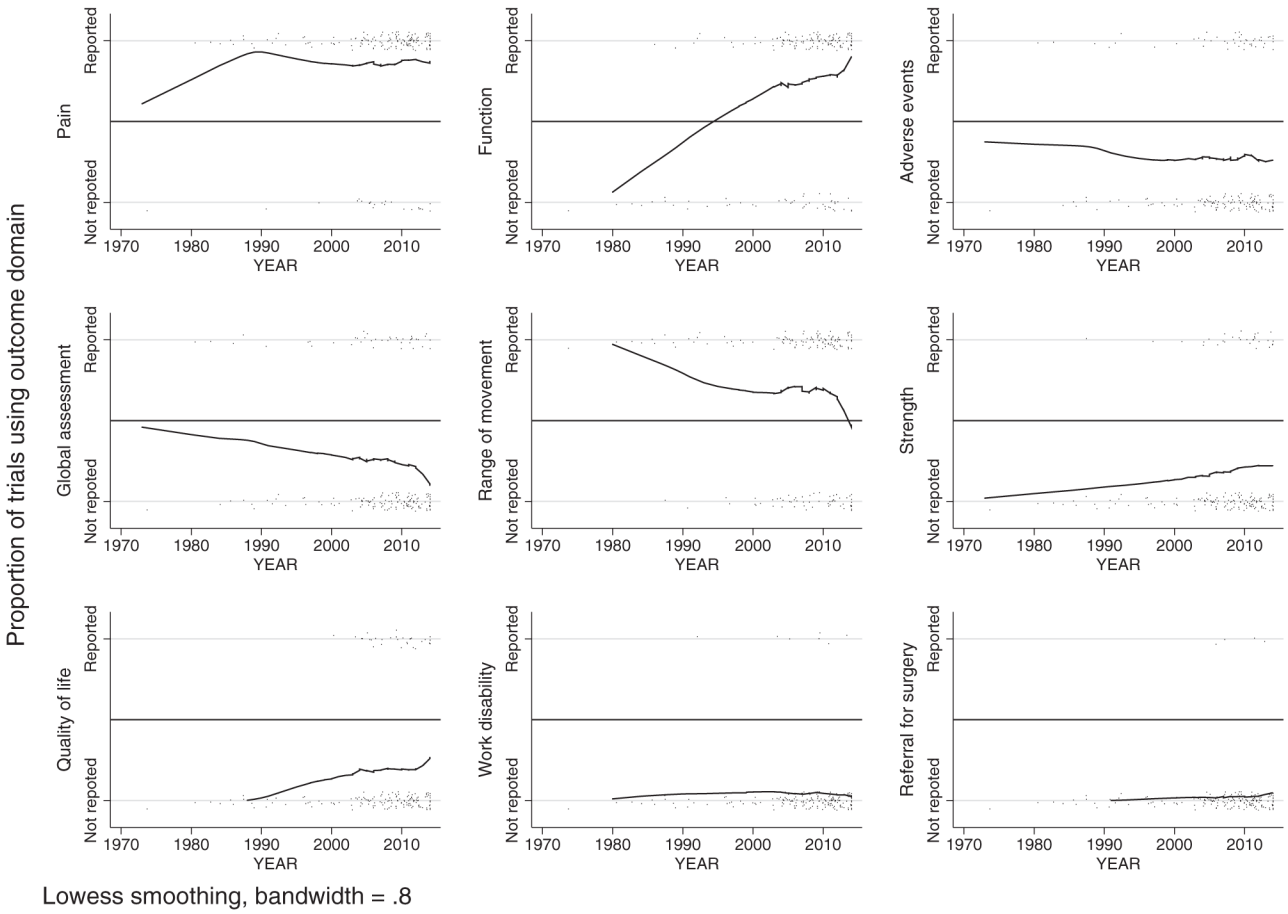

Fig. 2.

Smoothed scatterplots of outcome domain use over time. Solid lines are smoothed values calculated from the nonparametric regression method locally weighted scatterplot smoothing (lowess), using a bandwidth of 0.8. Smoothed values falling outside of range (<0 or >1) are not displayed. Observed values are binary (outcome reported or not reported) but are presented with some noise, so that individual observations are distinguishable (black points).

Table 1

Characteristics of included trials

Characteristics	Number (%)of trials, *n* = 171
Year of publication	
1973–2001	29 (17)
2002–2006	32 (19)
2007–2010	47 (27)
2011–2014	63 (37)
Country	
Turkey	33 (19)
United Kingdom	21 (12)
United States of America	15 (9)
India	12 (7)
Australia	10 (6)
Italy	10 (6)
Other (n < 10 trials per country)[a]	70 (41)
Shoulder pain diagnosis	
Rotator cuff disease	101 (59)
Adhesive capsulitis (frozen shoulder)	51 (30)
Nonspecific shoulder pain	19 (11)
Physical therapy intervention[b]	
Manual therapy alone	21 (12)
Exercise alone	25 (15)
Electrotherapy modality alone	24 (14)
Multimodal physical therapy intervention[c]	17 (10)
Manual therapy delivered as an add-on to another physical therapy intervention	36 (21)
Exercise delivered as an add-on to another physical therapy intervention	17 (10)
Electrotherapy modality delivered as an add-on to another physical therapy intervention	44 (26)
Multimodal physical therapy intervention delivered as an add-on to another active intervention (eg, glucocorticoid injection)	8 (5)
Duration of intervention in weeks, median (IQR)	4 (3, 6)
Number of intervention arms	
Two	131 (77)
Three	28 (16)
Four	9 (5)
Five	3 (2)
Participant characteristics	
Sample size, median (IQR)	49 (30, 74)
Mean age, median (IQR)	53 (49, 56)
Mean duration of condition (months), median (IQR)	7 (5, 14)
Percentage male, median (IQR)	41 (31, 52)

Abbreviations: IQR, interquartile range.

All values are given as *n* (%) except where indicated.

[a]Other countries include Austria, Belgium, Brazil, Canada, China, Cuba, Denmark, Egypt, Finland, France, Germany, Greece, Hong Kong, Iran, Japan, Kuwait, Mexico, Norway, Poland, Republic of Korea, Saudi Arabia, Serbia, Singapore, South Africa, Spain, Sweden, Taiwan, Thailand, The Netherlands.

[b]Percentages do not sum to 100 as some trials measured more than one type of intervention (eg, manual therapy alone versus exercise alone).

[c]Comprises either “manual therapy plus exercise,” “manual therapy plus electrotherapy,” “exercise plus electrotherapy,” or “manual therapy plus exercise plus electrotherapy.”

Table 2

Number (%) of trials reporting outcome domains and measurement instruments

Outcome domains and measurement instruments	Number (%) of trials, *n* = 171
Pain	148 (87)
Total number of measurement instruments = 35	
Overall pain (VAS)	69 (40)
Pain on movement (VAS)	38 (22)
Night pain (VAS)	32 (19)
Rest pain (VAS)	32 (19)
Shoulder Pain And Disability Index pain subscale [40]	30 (18)
Analgesic use recording form	12 (7)
Pressure pain threshold (algometry)	8 (5)
Day pain (VAS)	7 (4)
Overall pain (dichotomous measure)	7 (4)
Overall pain (NRS)	7 (4)
Constant–Murley pain subscale [41]	6 (4)
Other (instrument reported in ≤3% of trials)	45 (28)
Function/disability	123 (72)
Total number of measurement instruments = 29	
Constant–Murley score [41]	37 (22)
Shoulder Pain And Disability Index total scale [40]	32 (19)
Shoulder Pain And Disability Index disability subscale [40]	14 (8)
Dutch Shoulder Disability Questionnaire [42]	13 (8)
Disabilities of the Arm, Shoulder and Hand [43]	13 (8)
Function (VAS)	7 (4)
University of California-Los Angeles end-result score [44]	7 (4)
Function (categorical rating scale)	6 (4)
UK Shoulder Disability Questionnaire [45]	6 (4)
Other (instrument reported in ≤3% of trials)	34 (20)
Adverse events	47 (27)
Total number of measurement instruments = 1	
Patient-rated adverse event recording form	47 (27)
Global assessment of treatment success	41 (24)
Total number of measurement instruments = 4	
Patient-rated Likert scale	41 (24)
Clinician-rated Likert scale	9 (5)
Patient Global Impression of Change scale [46]	1 (1)
Patient Acceptable Symptom State	1 (1)
Range of movement	114 (67)
Total number of measurement instruments = 2	
Active range of movement via goniometer or tape measure (any movement, eg, abduction, flexion)	66 (39)
Passive range of movement via goniometer or tape measure (any movement, eg, abduction, flexion)	37 (22)

Outcome domains and measurement instruments	Number (%) of trials, *n* = 171
Range of movement via goniometer or tape measure (any movement, eg, abduction, flexion; unclear if active or passive)	37 (22)
Strength	31 (18)
Total number of measurement instruments = 6	
Isometric strength via dynamometer (any movement)	13 (8)
Isokinetic strength via dynamometer (any movement)	4 (2)
Other	17 (10)
Health-related quality of life	30 (18)
Total number of measurement instruments = 8	
Short-Form 36 Health Survey subscales (eg, (social functioning, vitality, general health) [47]	17 (10)
EuroQoL EQ-5D [48]	6 (4)
Other (instrument reported in ≤3% of trials)	10 (6)
Work disability	7 (4)
Total number of measurement instruments = 3	
Sick leave (days)	4 (2)
Return to work (days)	2 (1)
Occupational stress indicator [49]	1 (1)
Referral for surgery	4 (2)
Total number of measurement instruments = 1	
Self-reported or patient file review	4 (2)
Other	43 (25)
Total number of measurement instruments = 47	
Cost	7 (4)
Beck depression inventory (depression) [50]	6 (4)
Other[a] (instrument reported in ≤3% of trials)	30 (18)

Abbreviations: VAS, visual analog scale (0–10 or 0–100); NRS, numerical rating scale (0–10 or 0–100).

[a]Other measures further subclassified as presence of calcific deposits (measured using radiology), cost/health care use, physical examination maneuvers for rotator cuff disease (eg, Hawkins–Kennedy impingement test), morphologic features of the supraspinatus muscle (measured using radiography), muscle/tendon sensitivity, posture, proprioception, psychological symptoms (eg, depression), scapular dysfunction, severity of main complaint, shoulder stability, stiffness, supraspinatus tendon thickness (measured using ultrasound), swelling, weakness on movement.

Table 3

Specification of primary outcome domain and timing of outcome assessment

Characteristics	Number (%)of trials, *n* = 171
Specification of primary outcome domain	
At least one primary outcome domain specified	47 (27)
Frequency of primary outcome domains specified[a]	
Function/disability	28 (60)
Pain	9 (19)
Global assessment of treatment success	9 (19)
Range of movement	6 (13)
Presence of calcium deposits (measured using radiography)	1 (2)
Time points at which domains were measured	
Up to 3 wk	88 (51)
>3 wk and up to 6 wk	86 (50)
>6 wk and up to 6 mo	83 (49)
>6 mo	21 (12)
Time interval incorporated in the pain measurement instrument[b]	
Current pain	15 (10)
Average pain in the previous 24 hr	7 (5)
Average pain in the previous week	33 (22)
Average pain in the previous fortnight	1 (1)
Average pain in the previous month	1 (1)
Not specified	91 (61)

[a] Denominator = 47; however, some trials included more than one primary domain so percentages do not sum to 100%.

[b] Denominator = 148.

Table 4

Association between shoulder pain diagnosis and frequency of outcome domains

Outcome domain	Number (%)of RCD trials, *n* = 101	Number (%)of AC trials, *n* = 51	Risk ratio[a] (95% CI)
Pain	92 (91)	39 (76)	1.2 (1.01–1.4)
Function/disability	75 (74)	34 (67)	1.1 (0.9–1.4)
Adverse events	29 (29)	13 (25)	1.1 (0.6–2.0)
Global assessment of treatment success	21 (21)	9 (18)	1.2 (0.6–2.4)
Range of movement	59 (58)	42 (82)	0.7 (0.6–0.9)
Strength	26 (26)	1 (2)	13.1 (1.8–94.0)
Health-related quality of life	18 (18)	7 (14)	1.3 (0.6–2.9)
Work disability	7 (7)	0	8.2 (0.5–146.0)[b]
Referral for surgery	4 (4)	0	4.8 (0.3–90.0)[b]

Abbreviations: RCD, rotator cuff disease; AC, adhesive capsulitis.

Note that 19 trials including participants with nonspecific shoulder pain were excluded from the previously mentioned analysis.

[a]Risk ratio, risk of outcome reported in RCD trials/risk of outcome reported in AC trials.

[b]Odds ratios calculated from penalized likelihood logistic regression.