OPEN ACCESS

University of BRISTOL

Page, M., Huang, H., Verhagen, A. P., Gagnier, J. J., & Buchbinder, R. (2018). Outcome reporting in randomized trials for shoulder disorders: literature review to inform the development of a core outcome set. *Arthritis Care and Research*, *70*(2), 252-259. https://doi.org/10.1002/acr.23254

Peer reviewed version

Link to published version (if available):
10.1002/acr.23254

Link to publication record in Explore Bristol Research
PDF-document

This is the author accepted manuscript (AAM). The final published version (version of record) is available online via Wiley at http://onlinelibrary.wiley.com/doi/10.1002/acr.23254/abstract . Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
http://www.bristol.ac.uk/pure/about/ebr-terms

Outcome reporting in randomized trials for shoulder disorders: literature review to inform the development of a core outcome set

Matthew J Page, PhD, Hsiaomin Huang, MPH, Arianne P Verhagen, PhD, Joel J Gagnier, PhD[§], Rachelle Buchbinder, PhD[§]

[§] Contributed equally as senior authors of this work.

Corresponding author: Professor Rachelle Buchbinder, Monash Department of Clinical Epidemiology, Suite 41, Cabrini Medical Centre, 183 Wattletree Road, Malvern, Victoria, 3144, Australia. Phone: +61 3 9509 4445. Email: rachelle.buchbinder@monash.edu.

Funding: This project was supported by funding from a Patient Centered Outcomes Research Institute (PCORI) Eugene Washington Engagement Award #2072. MJP is supported by an Australian National Health and Medical Research Council (NHMRC) Early Career Fellowship (1088535). RB is supported by an Australian NHMRC Senior Principal Research Fellowship. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Word count: 3,199

ABSTRACT

Objectives

To explore the outcome domains and measurement instruments reported across randomized trials of any interventions for various shoulder disorders.

Methods

We searched for shoulder trials included in Cochrane reviews published up to Issue 10, 2015, or indexed in PubMed between 2006 and 2015. Trials were eligible for inclusion if they focused on any intervention for rotator cuff disease, adhesive capsulitis, shoulder instability, glenohumeral or acromioclavicular osteoarthritis, shoulder dislocation, proximal humeral or humeral head fractures, or unspecified shoulder pain. Two authors independently selected trials for inclusion and extracted information on the domains and measurement instruments reported, with consensus discussion among all authors where required.

Results

We included 409 trials, published between 1954 and 2015. Across the trials, we identified 319 different instruments that were classified into 32 domains. Most trials reported a measure of pain (90%), range of motion (78%), and physical function (71%). Measurement of adverse events was reported in only 31% of trials. Muscle strength was reported in 44% of trials and imaging outcomes were reported in 21% of trials. Other patient-reported outcome measures such as global assessment of treatment success, health-related quality of life, work ability, and psychological functioning, were each reported in 15% or fewer trials. Most domains were reported at a similar frequency across different shoulder disorders.

Conclusion

There was wide diversity in the domains and measurement instruments reported. Our results provide the foundation for the development of a core outcome set for use in future trials across all shoulder disorders.

SIGNIFICANCE AND INNOVATIONS

- Across 409 trials for various shoulder disorders (e.g. rotator cuff disease, adhesive capsulitis, shoulder instability, proximal humeral/humeral head fracture), we identified 319 measurement instruments that were classified into 32 domains – diversity so extensive that the ability to compare and synthesise the results of shoulder trials is severely hampered.
- The most common domains, consistently measured across all trials, were pain (90%), range of motion (78%) and physical function (71%), while other patient-reported outcome measures such as global assessment of treatment success, health-related quality of life, work ability, and psychological functioning, were each reported in 15% or fewer trials.
- Most domains were reported at a similar frequency across the different shoulder disorders, which suggests that it would be appropriate to develop a single core outcome set for all shoulder disorders.

Shoulder pain is a significant cause of morbidity and disability in the general population (1, 2), with an estimated point prevalence ranging from 7 to 26% (3). The most common cause of shoulder pain is rotator cuff disease, while less common causes include adhesive capsulitis, glenohumeral osteoarthritis, and instability or dislocations/fractures resulting from sport injuries in young adults (1, 4, 5). Although the causes of shoulder pain vary, there is strong commonality in endpoints, with most patients presenting with pain that disrupts sleep patterns, hampers the performance of daily activities such as dressing and bathing, and negatively impacts on recreation and work ability (6, 7). Further, the impact of persisting shoulder pain on earnings, missed workdays and disability payments, is substantial (8, 9).

Decisions makers often rely on the results of clinical trials to guide treatment decisions. However, it is not ideal if the outcomes that are measured have poor or unproven measurement properties, and if they vary across trials and hence cannot be compared or synthesized in meta-analyses (10-12). A promising strategy to reduce variation in outcome measurement is the development of core outcome sets (COSs) (12). This involves defining a set of outcome domains (constructs such as pain and function) that should be measured at a minimum in all trials for a particular condition, and then defining the measurement instruments that must be administered to cover a corresponding domain (13). COSs and their recommended measurement instruments have been developed for many musculoskeletal conditions (14), including rheumatoid arthritis (11, 15), low back pain (16) and gout (17), but not yet for shoulder disorders.

The lack of a COS for shoulder disorders may have contributed to the wide diversity in measurement observed in a review of 171 trials of physical therapy interventions for rotator cuff disease, adhesive capsulitis and unspecified shoulder pain (18). However, it is unclear

whether the diversity observed in that review is similar in trials for other shoulder disorders (e.g. shoulder instability, glenohumeral osteoarthritis) and for other interventions (e.g. glucocorticoid injections, tendon repair surgery). It is possible that other domains that are important to patients have been measured in trials of interventions other than physical therapy. Investigating the outcomes measured in a broader sample of shoulder trials should provide a stronger foundation for the development of a COS for use in future trials for all shoulder disorders.

The aim of this review was therefore to expand the previous review of physical therapy trials by exploring the frequency of outcome domains and measurement instruments reported across randomized trials of all interventions for a wider array of shoulder disorders. We also investigated whether reporting of domains varied by disorder.

MATERIALS AND METHODS

The methods for our literature review were pre-specified in a study protocol (19).

Eligibility criteria

We included randomized and quasi-randomized controlled trials investigating the effects of any intervention for one of the following shoulder disorders: rotator cuff disease (an umbrella term to classify disorders of the rotator cuff, including subacromial impingement syndrome, rotator cuff tendinopathy or tendinitis, partial or full rotator cuff tear, calcific tendinitis and subacromial bursitis (20)), adhesive capsulitis, shoulder instability, glenohumeral or acromioclavicular osteoarthritis, dislocation of the shoulder, proximal humeral or humeral head fractures, or unspecified shoulder pain. Since criteria used to diagnose shoulder disorders is not uniform across previous trials (4, 21), we included trials if they used any of

the labels specified above, rather than basing inclusion on specific diagnostic criteria. We excluded trials that enrolled patients with systemic inflammatory conditions such as rheumatoid arthritis, hemiplegia causing secondary shoulder pain, or pain in the shoulder region as part of a complex myofascial neck/shoulder/arm pain condition (e.g. complex regional pain syndrome). Trials were eligible if they compared any active intervention (e.g. manual therapy, surgery, glucocorticoid injection) to placebo, no treatment, or another active intervention. We only included completed trials, which were published and written in English.

Search methods

We used three approaches to identify trials. We included all trials examined in the previous review of domains and instruments used in physical therapy trials (18). We searched the Cochrane Database of Systematic Reviews (up to Issue 10, 2015) and screened the list of included studies in all Cochrane reviews of interventions for shoulder disorders. We also searched PubMed (January 2006 to December 2015) to identify recently published trials that may not have been included in the Cochrane reviews. The full Boolean search strategies for both databases are presented in online supplementary file 1.

Selection of trials

One author (either MJP or HH) screened all titles and abstracts against the eligibility criteria. The full text of articles assumed to be eligible, or with uncertain eligibility, were retrieved and screened independently by both authors. Discrepancies were resolved through discussion.

Data extraction and management

Data from Cochrane reviews and the PubMed-indexed trials were extracted by one author (either MJP or HH) and verified by another. Discrepancies were resolved through discussion until consensus was reached. Data were extracted from the 'Characteristics of included studies table' of each Cochrane review (which include detailed information on the domains and instruments reported in each trial), or from the full text report of each PubMed-indexed trial. We recorded for each trial the year of publication, diagnostic label used, sample size, intervention under investigation (for head-to-head trials, we recorded the first intervention mentioned in the objectives of the trial), and all outcome measurement instruments described either in the Methods or Results section of the paper. As defined by the Outcome Measures in Rheumatology (OMERACT) Filter 2.0, a measurement instrument could be a single question, a questionnaire, a score obtained through physical examination, a laboratory measurement, or a score obtained through observation of an image (13). If results of trials had been reported in multiple journal articles, we extracted data from the trial as a unit by incorporating information from all corresponding articles. Once data extraction was complete, the dataset was combined with the dataset used in the previous review of physical therapy trials (18).

Classification of outcome measurement instruments into domains

One author (MJP or HH) classified each reported outcome measurement instrument under the domain that its developers originally designed it to address. If the instrument was unknown to the authors, classification was guided by retrieving the article that first described the content or measurement properties of the instrument. One author (MJP) then classified each domain under one of the four 'Areas' of the OMERACT Filter 2.0: Life Impact (e.g. quality of life, activities of daily living), Resource Use (e.g. health care visits and associated costs),

Pathophysiological Manifestations (e.g. changes in body function and structure that accompany a condition), and Mortality (13). The same author classified domains using the International Classification of Functioning, Disability and Health (ICF) conceptual model developed by the World Health Organization (22). Two authors with expertise in shoulder-specific measurement instruments (RB and JG) verified all classifications for appropriateness.

Statistical analyses

We summarised results using frequencies and percentages for binary outcomes, and medians and interquartile ranges (IQRs) for continuous outcomes. Frequency of outcome domains was calculated for the complete set of trials and for trials sub-grouped by shoulder disorder (namely, rotator cuff disease, adhesive capsulitis, shoulder instability, glenohumeral/acromioclavicular osteoarthritis, shoulder instability, and fractures/dislocations; the latter were grouped together given their similar etiology). Analyses were undertaken using the statistical package Stata version 13 (23).

RESULTS

The inclusion criteria were met by 121 trials included in 18 Cochrane reviews, and 117 PubMed-indexed trials, which were combined with the 171 physical therapy trials evaluated in the previous review (18). Therefore, we evaluated 409 unique trials in total (Figure 1). The trials were published between 1954 and 2015 (Table 1). Most trials were for rotator cuff disease (186/409 [45%]), followed by adhesive capsulitis (97/409 [24%]), and shoulder instability or proximal humeral/humeral head fracture (both in 31/409 [8%] of trials). There were 18 types of intervention under investigation across the trials. Manual therapy or exercise (delivered alone or in combination) was the most common intervention (140/409 [34%]),

followed by surgery (104/409 [25%]), electrotherapy modalities (such as therapeutic ultrasound) (65/409 [16%]), and glucocorticoid injection (45/409 [11%]). Trials included a median of 50 (IQR 38-75) participants.

Across the trials, we identified 319 different instruments that were classified into 32 domains. The majority of trials included a measure of pain (368/409 [90%]), range of motion (320/409 [78%]), and physical function (289/409 [71%]). Measurement of adverse events (serious or non-serious) was reported in only 126/409 (31%) trials. A measure of muscle strength was reported in 178/409 (44%) trials, and a radiographic outcome (i.e. any measure of the structure of the bones/joints/tendons evaluated via imaging) was reported in 86/409 (21%) trials. Other domains (e.g. global assessment of treatment success, health-related quality of life, work ability, health care service use, psychological functioning) were each reported in 15% or fewer trials. Assessment of the number of deaths was reported in only 13/409 (3%) trials.

Using the OMERACT Filter 2.0, 11 domains were classified under the Life Impact Area (including pain, physical function, global assessment of treatment success; Table 2). Fifteen domains were classified under the Pathophysiological Manifestations Area (including range of motion, muscle strength, and radiographic outcomes; Table 3). Four domains were classified under the Resource Use Area (e.g. health care services, work productivity; Table 4). Number of deaths and adverse events do not fall under these three Areas, but make up the last two domains. Using the ICF conceptual model, all of the domains in Table 2 fall under the Activities and Participation domains, while all of the domains in Table 3 fall under the Body Functions and Body Structures domains.

A wide range of instruments was used to measure the same outcome across the included trials (online supplementary Table S1). Pain was measured by 47 different instruments, which varied in terms of the type of pain addressed (e.g. night pain intensity, pain at rest or with activity), descriptor for the maximum score on the scale (e.g. "intolerable pain", "worst pain imaginable"), and period of interest (e.g. "pain within the last 24 hours" versus "pain within the last week"). The most common was overall pain measured using a visual analogue scale (VAS) (174/409 trials [43%]). There were 49 different measures of physical function, and of these, the Constant-Murley score (24) was used most often (128/409 trials [31%]), followed by the Shoulder Pain And Disability Index (SPADI) (25) (49/409 trials [12%]). Nearly all (44/49 [90%]) of these instruments were patient-reported outcome measures (PROMs), while four instruments (including the Constant-Murley score and ASES score) include physician-observed measures of functional impairment. Radiography was used to measure various structural outcomes, such as rotator cuff repair integrity, subluxation (partial dislocation), acromiohumeral distance and presence of calcific deposits. Of the measures of muscle strength, most were performed using a dynamometer to measure isokinetic or isometric strength in different positions (e.g. strength in flexion, strength in abduction). Range of motion was measured using either a goniometer or tape measure, although there was variation in the type of movements assessed (e.g. flexion, extension, abduction), the number of movements assessed, and whether movements were active or passive. Other domains were each assessed by a median of three instruments (range 1-12).

Most domains (24/32 [75%]), including pain, physical function and range of motion, were reported at a similar frequency across the different shoulder disorders (Figure 2, online supplementary Table S2). Exceptions to this included: trials for adhesive capsulitis less frequently reported a measure of muscle strength; trials for shoulder fracture/dislocation more

frequently reported a measure of adverse events, number of deaths, radiographic outcomes, requiring re-operation and failure of surgery; trials for osteoarthritis more frequently included a measure of radiographic outcomes; and trials for shoulder instability more frequently included a measure of instability and recreation/leisure activities.

DISCUSSION

We found that the outcome domains and measurement instruments in trials of many different interventions for several shoulder disorders are widely diverse. Across the trials we identified 319 different instruments which were classified into 32 domains. Most trials included a measure of pain (90%), range of motion (78%), and physical function (71%). Measurement of muscle strength was reported in 44% of trials and measurement of radiographic outcomes was reported in 21% of trials. Other domains, particularly patient-reported outcome measures (PROMs) such as global assessment of treatment success, health-related quality of life, work ability, and psychological functioning, were each reported in 15% or fewer trials. There were 47 different measurement instruments for pain and 49 for function across the trials; nearly all of the latter were PROMs. Most domains were reported at a similar frequency across the different shoulder disorders. Measurement of adverse events was reported in only 31% of trials.

Strengths and limitations

A strength of our study is the inclusion of trials investigating a diverse set of interventions for an array of shoulder disorders, which enhances the generalisability of the findings. In addition, the majority of our sample of trials was identified from published Cochrane reviews, each of which used a comprehensive search strategy and methods to minimise error in trial selection. However, some limitations need to be considered. Selective reporting of

only the positive/statistically significant outcomes is common in clinical trials (26), and so by relying on what was reported rather than asking trialists about any non-reported domains and instruments, we may have underestimated the frequency of outcome domains and instruments that were *actually* measured. Also, rather than updating the searches of each Cochrane review on a shoulder disorder, we only searched PubMed for shoulder trials indexed in the last 10 years. Therefore, it is likely that we have not identified *all* published trials for shoulder disorders. However, we are not aware of any evidence that outcomes measured in trials included in Cochrane reviews or indexed in PubMed differ to those in trials indexed elsewhere. Therefore, we do not believe our findings are unrepresentative of all shoulder trials.

Comparison with other studies

This review extends the findings of a previous review of physical therapy trials (18) in several ways. We included a much larger sample of trials (409 versus 171) addressing a more comprehensive set of questions. This led to the identification of 12 domains that were not previously recorded, namely: number of deaths, radiographic outcomes, shoulder instability, failure of surgery, surgical process outcomes, requiring re-operation or revision surgery, satisfaction with treatment services, social functioning, recreation or leisure activity, sleep functioning, non-health care service use, and haemodynamic variables. We also identified a larger number of measurement instruments for the same domain (for example, 35 pain instruments and 29 function instruments were noted in the previous review, compared with 47 and 49, respectively, in the current sample). Using the OMERACT Filter 2.0 (13) and ICF framework (22) to classify domains, we found that slightly more (15/32 versus 11/32) of the domains used in shoulder trials were measures of changes in body function and structure that accompany the disorder (that is, “Pathophysiological Manifestations” in OMERACT terms or

"Body Functions and Structures" in the ICF conceptual model) rather than measures of performance of activities of daily living, work and recreation/leisure ("Life Impact" in OMERACT terms or "Activities and Participation" in the ICF conceptual model). Finally, measurement of adverse events was reported only slightly more often in the current sample (31% versus 27%), suggesting that trials of interventions other than physical therapy are not immune from poor harms reporting.

Explanation and implications of study results

Most of the domains (75%) we identified were measured at a similar frequency across all of the included shoulder disorders (including pain, physical function, and range of motion). For this reason, we think it will be appropriate to develop a single COS for all shoulder disorders, rather than developing different sets for different disorders (i.e. one for rotator cuff disease, one for adhesive capsulitis). A single COS would not preclude trialists from measuring other domains that might be especially relevant to the context of the trial (13, 27). For example, a trial of surgical repair for rotator cuff tear may use imaging to measure integrity of the repair, while a trial of an intervention designed to improve shoulder posture may measure improvement in posture, and a trial for shoulder instability in young athletes may measure time to return to sport. However, we recognise that the decision to create a single COS for all shoulder disorders requires consideration from various stakeholders, the most important of whom is patients, who can help ensure that the COS sufficiently captures the experience of their condition (28).

Government-sponsorship of initiatives such as the Patient-Centered Outcomes Research Institute (PCORI) indicates that policy makers have a strong desire for PROMs to be collected in clinical trials (29). Previous shoulder trials have done well at measuring patient-

reported pain, but infrequently measured other potentially important PROMs such as global assessment of treatment success, health-related quality of life, work ability, sleep quality and psychological functioning. Also, while many patient-reported measures of physical functioning were identified across the trials, the most commonly used one – the Constant-Murley score, used in 31% of trials – includes physician-rated components to measure range of motion and strength (impairment measures). It is unclear if trialists believe pain is the only necessary PROM to measure in shoulder trials, have concerns about the measurement instruments available to measure other PROMs, or believe that non-PROMs such as range of motion (measured using a goniometer) or radiographic outcomes are more valid and reliable. Many studies have found that PROMs do not correlate well with objectively measured outcomes in people with shoulder disorders (30-33), yet provide valuable insight into the overall burden that a health condition places on an individual (34). In future, it would be valuable to ask different stakeholders for their opinion on the relative value of PROMs and non-PROMs for shoulder disorders, and to consider which PROMs are the most essential to include in a COS for shoulder disorders.

Additional research is needed before we can recommend a COS for shoulder disorders, and the results of this literature review should therefore be considered preliminary only. It is possible that the outcome domains previously measured in shoulder trials do not fully represent the lived experience of people with shoulder disorders. Other domains that have never been measured in past trials may exist, and could be elicited from patients using qualitative methods, such as focus groups and interviews (35). Further, a plethora of instruments have been collated in this review, but we have not yet evaluated their measurement properties, including construct validity, test-retest reliability, and responsiveness. The instruments most frequently used in past trials may not necessarily be the

best at reliably detecting change in symptoms/function over time. A systematic review of measurement properties of instruments for shoulder disorders will help determine which are most fit for purpose.

In conclusion, there is a wide variety of domains and instruments being used in trials for various shoulder disorders. The most common domains, consistently measured across all trials, were pain, range of motion and physical function. We will use the results of this review to inform an international Delphi study to select core domains, and then review the measurement properties of applicable instruments. Such research will ultimately lead to the creation of a core set of domains and recommended instruments for use in trials for shoulder disorders, whose endorsement we will seek from OMERACT.

Contributors

JJG and RB conceived the study design. MJP and HH provided input on the study design. MJP and HH extracted data. MJP and HH classified outcome measurement instruments into domains, and JJG and RB verified that all classifications were appropriate MJP undertook the statistical analyses. MJP wrote the first draft of the manuscript. All authors contributed to revisions of the manuscript. All authors approved the final version of the submitted manuscript.

Competing interest

RB is an author of four trials included in this review, and of nine of the Cochrane reviews that were used to identify trials. However, RB was not involved in the eligibility assessment or data extraction. All other authors declare no competing interests.

REFERENCES

1. 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.
2. Ostor AJ, Richards CA, Prevost AT, Speed CA, Hazleman BL. Diagnosis and relation to general health of shoulder disorders presenting to primary care. Rheumatology (Oxford). 2005;44(6):800-5.
3. 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. Scandinavian Journal of Rheumatology. 2004;33:73-81.
4. 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.
5. Tekavec E, Jöud E, Rittner R, Mikoczy Z, Nordander C, Petersson IF, et al. Population-based consultation patterns in patients with shoulder pain diagnoses. BMC Musculoskeletal Disorders. 2012;13:238.
6. van der Heijden GJ. Shoulder disorders: a state-of-the-art review. Baillieres Best Pract Res Clin Rheumatol. 1999;13:287-309.
7. Largacha M, Parsons Iv IM, Campbell B, Titelman RM, Smith KL, Matsen Iii F. Deficits in shoulder function and general health associated with sixteen common shoulder diagnoses: A study of 2674 patients. Journal of Shoulder and Elbow Surgery. 2006;15(1):30-9.
8. Mather RC, 3rd, Koenig L, Acevedo D, Dall TM, Gallo P, Romeo A, et al. The societal and economic value of rotator cuff repair. J Bone Joint Surg Am. 2013;95(22):1993-2000.

9. 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 Musculoskeletal Disorders. 2012;13(1):17.

10. Clarke M. Standardising outcomes for clinical trials and systematic reviews. Trials. 2007;8:39.

11. 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(1):38.

12. 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. Journal of Health Services Research & Policy. 2012;17(1):1-2.

13. 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. Journal of Clinical Epidemiology. 2014;67(7):745-53.

14. Gorst SL, Gargon E, Clarke M, Blazeby JM, Altman DG, Williamson PR. Choosing Important Health Outcomes for Comparative Effectiveness Research: An Updated Review and User Survey. PLoS One. 2016;11(1):e0146444.

15. Boers M, Tugwell P, Felson DT, van Riel PL, Kirwan JR, Edmonds JP, et al. World Health Organization and International League of Associations for Rheumatology core endpoints for symptom modifying antirheumatic drugs in rheumatoid arthritis clinical trials. Journal of Rheumatology. 1994;S41(21):86-9.

16. 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. European Spine Journal. 2015:1-16.

17. Schumacher HR Jr, Edwards LN, Perez-Ruiz F, Becker M, Chen LX, Furst DE, et al. Outcome measures for acute and chronic gout. Journal of Rheumatology. 2005;32(12):2452-5.

18. Page MJ, McKenzie JE, Green SE, Beaton DE, Jain NB, Lenza M, et al. Core domain and outcome measurement sets for shoulder pain trials are needed: systematic review of physical therapy trials. J Clin Epidemiol. 2015;68(11):1270-81.

19. Gagnier JJ, Page MJ, Huang H, Verhagen AP, Buchbinder R. Creation of a core outcome set for clinical trials of patients with shoulder pain: A protocol. Trials. 2016:In press.

20. Whittle S, Buchbinder R. In the clinic. Rotator cuff disease. Ann Intern Med. 2015;162(1):Itc1-15.

21. Schellingerhout JM, Verhagen AP, Thomas S, Koes BW. Lack of uniformity in diagnostic labeling of shoulder pain: time for a different approach. Manual Therapy. 2008;13:478-83.

22. World Health Organization. International Classification of Functioning, Disability and Health (ICF). Geneva: World Health Organization; 2001.

23. StataCorp. Stata Statistical Software: Release 13. College Station, TX: StataCorp LP; 2013.

24. Constant CR, Murley AH. A clinical method of functional assessment of the shoulder. Clinical orthopaedics and related research. 1987(214):160-4.

25. Roach KE, Budiman-Mak E, Songsiridej N, Lertratanakul Y. Development of a shoulder pain and disability index. Arthritis care and research : the official journal of the Arthritis Health Professions Association. 1991;4(4):143-9.

26. 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.

27. Williamson P, Altman D, Blazeby J, Clarke M, Devane D, Gargon E, et al. Developing core outcome sets for clinical trials: issues to consider. Trials. 2012;13:132.

28. Kirwan JR, Hewlett SE, Heiberg T, Hughes RA, Carr M, Hehir M, et al. Incorporating the patient perspective into outcome assessment in rheumatoid arthritis--progress at OMERACT 7. The Journal of rheumatology. 2005;32(11):2250-6.

29. Tunis SR, Clarke M, Gorst SL, Gargon E, Blazeby JM, Altman DG, et al. Improving the relevance and consistency of outcomes in comparative effectiveness research. Journal of comparative effectiveness research. 2016;5(2):193-205.

30. Russell RD, Knight JR, Mulligan E, Khazzam MS. Structural integrity after rotator cuff repair does not correlate with patient function and pain: a meta-analysis. J Bone Joint Surg Am. 2014;96(4):265-71.

31. Wylie JD, Suter T, Potter MQ, Granger EK, Tashjian RZ. Mental Health Has a Stronger Association with Patient-Reported Shoulder Pain and Function Than Tear Size in Patients with Full-Thickness Rotator Cuff Tears. J Bone Joint Surg Am. 2016;98(4):251-6.

32. Dunn WR, Kuhn JE, Sanders R, An Q, Baumgarten KM, Bishop JY, et al. Symptoms of pain do not correlate with rotator cuff tear severity: a cross-sectional study of 393 patients with a symptomatic atraumatic full-thickness rotator cuff tear. J Bone Joint Surg Am. 2014;96(10):793-800.

33. Curry EJ, Matzkin EE, Dong Y, Higgins LD, Katz JN, Jain NB. Structural Characteristics Are Not Associated With Pain and Function in Rotator Cuff Tears:

The ROW Cohort Study. Orthopaedic journal of sports medicine. 2015;3(5):2325967115584596.

34. Black N. Patient reported outcome measures could help transform healthcare. BMJ. 2013;346:f167.

35. Keeley T, Williamson P, Callery P, Jones LL, Mathers J, Jones J, et al. The use of qualitative methods to inform Delphi surveys in core outcome set development. Trials. 2016;17(1):230.

TABLES

Table 1. Characteristics of included trials

Characteristic	Number (%), of 409 trials
Year of publication	
1954– 2001	90 (22)
2002 – 2006	55 (13)
2007 – 2010	103 (25)
2011 - 2015	161 (39)
Shoulder disorder	
Rotator cuff disease	186 (45)
Adhesive capsulitis (frozen shoulder)	97 (24)
Shoulder instability	31 (8)
Proximal humeral or humeral head fracture	31 (8)
Unspecified shoulder pain	30 (7)
Dislocation of the shoulder	17 (4)
Glenohumeral or acromioclavicular osteoarthritis	16 (4)
Mixed (some with rotator cuff disease, others with instability)	1 (0.2)
Intervention under investigation	
Manual therapy or exercise (or both delivered in combination)	140 (34)
Surgery (e.g. tendon repair surgery for rotator cuff tear)	104 (25)
Electrotherapy modalities (e.g. therapeutic ultrasound, laser therapy)	65 (16)
Glucocorticoid injection	45 (11)
Platelet-rich plasma therapies	11 (3)
Acupuncture	8 (2)

Oral glucocorticoids	8 (2)
Intra-articular lignocaine	5 (1)
Arthrographic distension of the glenohumeral joint with glucocorticoid and saline	4 (1)
Sodium hyaluronate injection	4 (1)
Topical glyceryl trinitrate	3 (1)
Botulinum toxin	3 (1)
Extracorporeal shockwave therapy	3 (1)
Continuous interscalene brachial plexus block	2 (0.5)
Manipulation under anaesthesia	1 (0.2)
Needling fragmentation irrigation	1 (0.2)
Radiotherapy	1 (0.2)
Suprascapular nerve block	1 (0.2)

Table 2. OMERACT "Life Impact" (ICF "Activities and Participation") outcome domains in 409 included trials

Domain (examples in parentheses)	Number (%)	Number of instruments
Pain (how much a person's shoulder hurts overall, during or following activity, at night, etc.)	368 (90)	47
Physical functioning (ability to carry out physical activities of daily living such as dressing, bathing)	289 (71)	49
Global assessment of treatment success (person's assessment of their recovery or degree of improvement)	62 (15)	9
Health-related quality of life (physical, psychological and social domains of health)	62 (15)	12
Satisfaction with treatment services (person's satisfaction with care received)	61 (15)	6
Social functioning (ability to engage in normal social activities with family, friends)	29 (7)	2
Work ability (ability to meet physical or psychological demands of work)	24 (6)	4
Recreation and leisure activity (ability to engage in recreational or leisure activities, including sports)	22 (5)	7
Psychological functioning (depression, anxiety)	18 (4)	10
Severity of the main complaint (how much a person's main concern, be it pain or disability, bothers them)	5 (1)	1
Sleep functioning (impact of disorder on onset, maintenance, quality and amount of sleep)	2 (0.5)	2

Table 3. OMERACT "Pathophysiological Manifestations" (ICF "Body Functions and Structures") outcome domains in 409 included trials

Domain (examples in parentheses)	Number (%)	Number of instruments
Range of motion (distance and direction that the shoulder is able to move, for example, in flexion or abduction, either when initiated by the person or guided by the care provider)	320 (78)	25
Muscle strength (force generated by the contraction of a shoulder muscle)	178 (44)	38
Radiographic outcomes (any measure of the structure of the bones/joints/tendons measured via imaging, such as rotator cuff repair integrity, acromiohumeral distance)	86 (21)	46
Shoulder instability (loosening of connective tissue surrounding the shoulder joint, which can sometimes result in the head of the upper arm bone slipping out of the shoulder socket)	25 (6)	3
Failure of surgery (inability of the surgery to correct a specific structural problem, such as malunion of a fracture after surgery)	24 (6)	8
Testing positive on specific tests during physical examination (e.g. painful arc test for rotator cuff disease, relocation test for shoulder instability)	21 (5)	8
Surgical process outcomes (measures of the implementation of a surgical intervention)	20 (5)	7
Pain on palpation (pain following touch of any muscle,	7 (2)	1

Domain (examples in parentheses)	Number (%)	Number of instruments
tendon and bony prominences of the shoulder region)		
Scapular dysfunction (problems with rotating the scapular or poor motor control of the scapular)	7 (2)	8
Proprioception of the shoulder (person's sense of position and movement of the shoulder)	5 (1)	2
Muscle tone (slight tension or firmness present in resting muscles of the shoulder)	4 (1)	7
Shoulder posture (position in which shoulder is held upright against gravity while standing, sitting or lying down)	2 (0.5)	1
Weakness on movement (person's sense that more effort than normal is required to move the shoulder)	1 (0.2)	1
Shoulder swelling (abnormal enlargement caused by accumulation of fluid in the shoulder joint)	1 (0.2)	1
Haemodynamic variables (any measure of blood flow, such as blood pressure or heart rate)	1 (0.2)	1

Table 4. OMERACT "Resource Use" outcome domains in 409 included trials

Domain (examples in parentheses)	Number (%)	Number of instruments
Requiring re-operation or revision surgery (use of additional surgical procedures following failure of the first attempt)	35 (9)	1
Health care services (health care visits, laboratory tests, imaging, days of admission to a hospital, medications)	29 (7)	4
Non-health care services (visits to professionals of alternative medicine, over-the-counter medications, patient's time and travel expenses)	29 (7)	4
Work productivity (economic impact of absenteeism due to shoulder disorder)	23 (6)	2

FIGURE LEGENDS

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

SR = systematic review

Figure 2. Percentage of trials measuring the most commonly measured domains, sub-grouped by shoulder disorder. Values are percentages. AC = adhesive capsulitis, HRQoL = health-related quality of life, RCD = rotator cuff disease OA = osteoarthritis