

Outcome Reporting in Randomized Trials for Shoulder Disorders: Literature Review to Inform the Development of a Core Outcome Set

MATTHEW J. PAGE,¹ HSIAOMIN HUANG,² ARIANNE P. VERHAGEN,³ JOEL J. GAGNIER,² AND RACHELLE BUCHBINDER⁴

Objective. 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%). The recording of adverse events was reported in only 31% of the trials. Muscle strength was reported in 44% of the trials and imaging outcomes in 21%. 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% of the trials. Most of the domains were reported at similar frequencies across different shoulder disorders.

Conclusion. The domains and measurement instruments reported were widely diverse. Our results provide a foundation for the development of a core outcome set for use in future trials across all shoulder disorders.

INTRODUCTION

Shoulder pain is a significant cause of morbidity and disability in the general population (1,2), with an estimated point prevalence of 7–26% (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 sports-related injuries in young adults (1,4,5). Although the causes of shoulder pain vary, there is strong commonality in end points, with most patients presenting with pain that disrupts sleep patterns, hampers the performance of daily activities such as dressing and bathing, and negatively affects

recreation and work ability (6,7). Further, the impact of persistent shoulder pain in terms of 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 (12). This involves defining a set of outcome domains (i.e., constructs such as pain and function) that should be measured at a minimum in all trials for a particular

Supported by a Patient-Centered Outcomes Research Institute Eugene Washington Engagement Award (2072). Dr. Page is supported by an Australian National Health and Medical Research Council Early Career Fellowship (1088535). Dr. Buchbinder is supported by an Australian National Health and Medical Research Council Senior Principal Research Fellowship.

¹Matthew J. Page, PhD: Monash University, Melbourne, Victoria, Australia; ²Hsiaomin Huang, MPH, Joel J. Gagnier, PhD: University of Michigan, Ann Arbor; ³Arianne P. Verhagen, PhD: Erasmus University Medical Centre, Rotterdam, The Netherlands; ⁴Rachelle Buchbinder, PhD: Cabrini Medical Centre, Monash University, Malvern, Victoria, Australia.

Because Dr. Buchbinder was an author of 4 of the trials included in the review and 9 of the Cochrane reviews that were used to identify the trials, she was not involved in the eligibility assessment or data extraction.

Drs. Gagnier and Buchbinder contributed equally to this work.

Address correspondence to Rachelle Buchbinder, PhD, Monash Department of Clinical Epidemiology, Suite 41, Cabrini Medical Centre, 183 Wattletree Road, Malvern, Victoria, 3144, Australia. E-mail: rachelle.buchbinder@monash.edu.

Submitted for publication November 29, 2016; accepted in revised form April 4, 2017.

Significance & Innovations

- Across 409 trials for various shoulder disorders (e.g., rotator cuff disease, adhesive capsulitis, shoulder instability, and proximal humeral/humeral head fracture), we identified 319 measurement instruments that were classified into 32 domains. This diversity is so extensive that the ability to compare and synthesize 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 $\leq 15\%$ of the trials.
- Most of the domains were reported at similar frequencies across the different shoulder disorders, which suggests that it would be appropriate to develop a single core outcome set for all shoulder disorders.

condition, and then defining the measurement instruments that must be administered to cover a corresponding domain (13). Core outcome sets 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 core outcome set 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 or glenohumeral osteoarthritis) and for other interventions (e.g., glucocorticoid injections or 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 core outcome set 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 the reporting of domains varied by disorder.

MATERIALS AND METHODS

Eligibility criteria. The methods for our literature review were prespecified in a study protocol (19). 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 the criteria used to diagnose shoulder disorders is not uniform across 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, or glucocorticoid injection) to placebo, no treatment, or another active intervention. We only included completed trials that were written and published in English.

Search methods. We used 3 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 shown in Supplementary Appendix A (see Supplementary Appendix A, available on the *Arthritis Care & Research* web site at <http://onlinelibrary.wiley.com/doi/10.1002/acr.23254/abstract>).

Selection of trials. One of 2 authors (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 1 author (either MJP or HH) and verified by the second author. Discrepancies were resolved through discussion until consensus was reached. Data were extracted from the characteristics of included studies table of each Cochrane review (which includes 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 trial objectives), and all outcome measurement instruments described either in the Methods or Results section of the article. 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 data set was combined with the data set used in the previous review of physical therapy trials (18).

Classification of outcome measurement instruments into domains. One of 2 authors (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 1 of the 4 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], pathophysiologic 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 (JJG and RB) verified all classifications for appropriateness.

Statistical analysis. We summarized results using frequencies and percentages for binary outcomes and medians and interquartile ranges (IQRs) for continuous outcomes. The frequency of outcome domains was calculated for the complete set of trials and for trials subgrouped by shoulder disorder (namely, rotator cuff disease, adhesive capsulitis, shoulder instability, glenohumeral/acromioclavicular osteoarthritis, shoulder instability, and fractures/dislocations; fractures and dislocations were grouped together because of their similar etiology). Analyses were undertaken using 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

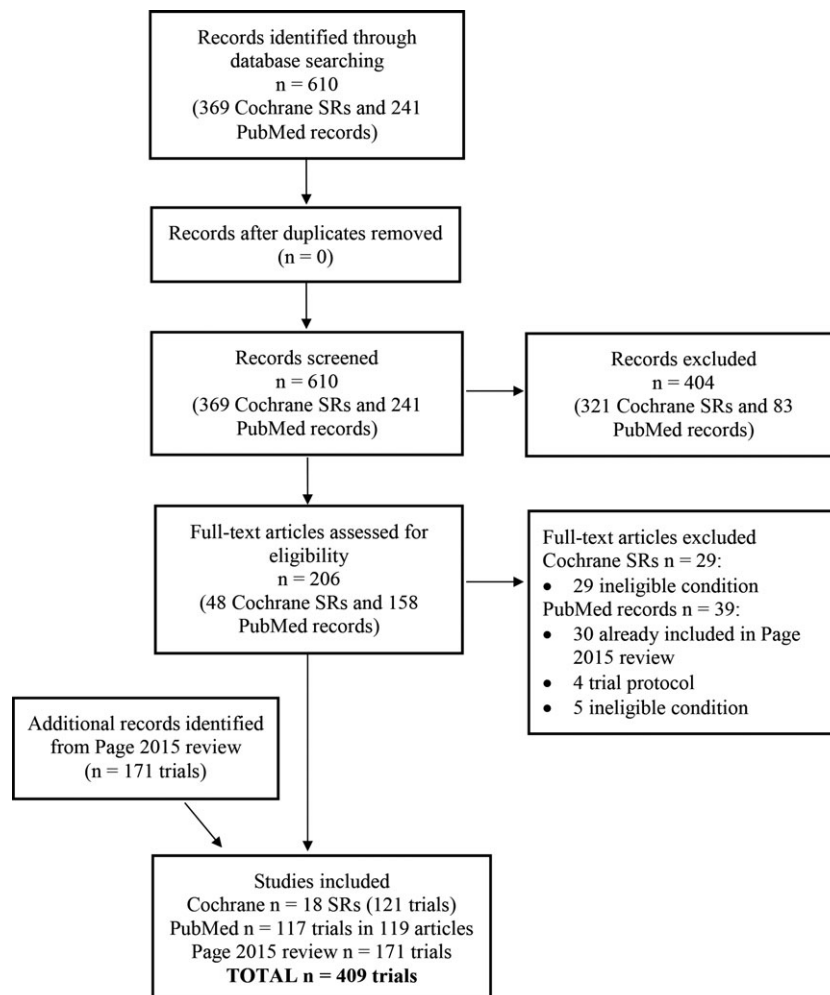


Figure 1. Flow diagram of the process of identifying and screening the trials included in this review. SRs = systematic reviews.

Table 1. Characteristics of the 409 included trials

Characteristic	No. (%)
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)
Intraarticular 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 anesthesia	1 (0.2)
Needling fragmentation irrigation	1 (0.2)
Radiotherapy	1 (0.2)
Suprascapular nerve block	1 (0.2)

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 trials [8%]). There were 18 types of interventions 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%). The trials included a median of 50 (IQR 38–75) participants.

Across the trials, we identified 319 different instruments classified into 32 domains. The majority of the trials included a measure of pain (368/409, 90%), range of motion (320/409, 78%), and physical function (289/409, 71%). The measurement of adverse events (serious or nonserious) was reported in only 126/409 trials (31%). A measure of muscle strength was reported in 178/409 trials (44%), 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 services use, or psychological functioning) were each reported in $\leq 15\%$ of the trials. Assessment of the number of deaths was reported in only 13/409 trials (3%).

Using the OMERACT filter 2.0, 11 domains were classified under the life impact area (including pain, physical function, and global assessment of treatment success) (Table 2). Fifteen domains were classified under the pathophysiologic 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 or work productivity) (Table 4). The numbers of deaths and adverse events do not fall under these three areas, but make up the last 2 domains. Using the ICF conceptual model, all of the domains shown in Table 2 fall under activities and participation, while all of the domains shown in Table 3 fall under body functions and body structures.

Table 2. Outcome Measures in Rheumatology life impact (International Classification of Functioning, Disability, and Health activities and participation) outcome domains in the 409 included trials

Domain (definition/examples)	No. (%)	No. 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 or 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. Outcome Measures in Rheumatology pathophysiologic manifestations (International Classification of Functioning, Disability, and Health body functions and structures) outcome domains in the 409 included trials

Domain (definition/examples)	No. (%)	No. 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, tendon, and bony prominences of the shoulder region)	7 (2)	1
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
Hemodynamic variables (any measure of blood flow, such as blood pressure or heart rate)	1 (0.2)	1

A wide range of instruments was used to measure the same outcome across the included trials (see Supplementary Table 1, available on the *Arthritis Care & Research* web site at <http://onlinelibrary.wiley.com/doi/10.1002/acr.23254/abstract>). Pain was measured by 47 different instruments, which varied in terms of the type of pain addressed (e.g., night pain intensity and pain at rest or with activity), descriptor for the maximum score on the scale (e.g., "intolerable pain" or "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 analog 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 (25) (49/409 trials [12%]). Nearly all (44/49 [90%]) of these instruments were patient-reported outcome measures (PROMs), and 4 instruments

(including the Constant-Murley score and the American Shoulder and Elbow Surgeons 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 or 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, or abduction), the number of movements assessed, and whether movements were active or passive. Other domains were assessed by a median of 3 instruments (range 1–12).

Most domains (24/32, 75%), including pain, physical function, and range of motion, were reported at similar

Table 4. Outcome Measures in Rheumatology resource use outcome domains in the 409 included trials

Domain (definition/examples)	No. (%)	No. instruments
Requiring reoperation 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

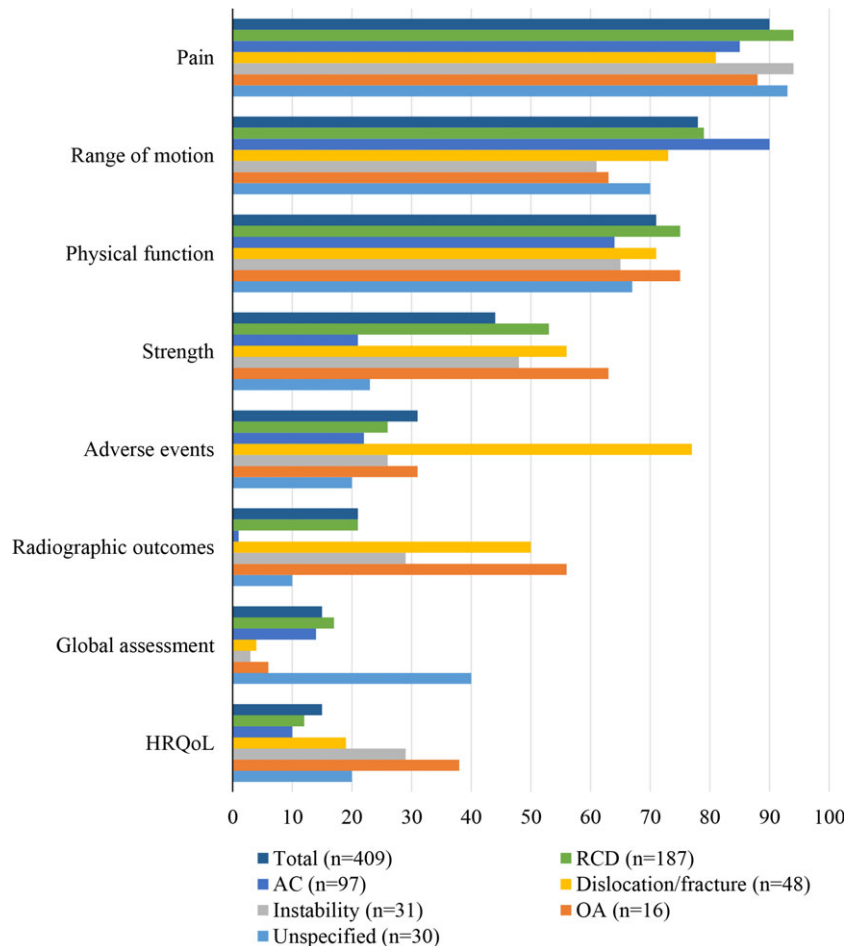


Figure 2. Percentage of trials measuring the most commonly measured domains, subgrouped by shoulder disorder. HRQoL = health-related quality of life; RCD = rotator cuff disease; AC = adhesive capsulitis; OA = osteoarthritis.

frequencies across the different shoulder disorders (Figure 2 and Supplementary Table 2, available on the *Arthritis Care & Research* web site at <http://onlinelibrary.wiley.com/doi/10.1002/acr.23254/abstract>). Exceptions to this included the following: 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 reoperation 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 the trials, and measurement of radiographic

outcomes was reported in 21% of the trials. Other domains, particularly PROMs such as global assessment of treatment success, health-related quality of life, work ability, and psychological functioning, were each reported in $\leq 15\%$ of the 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 the trials.

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 generalizability 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 minimize 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 nonreported 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 searched PubMed only 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 from those in trials indexed elsewhere. Therefore, we do not believe our findings are unrepresentative of all shoulder trials.

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 the number of deaths, radiographic outcomes, shoulder instability, failure of surgery, surgical process outcomes, requiring reoperation or revision surgery, satisfaction with treatment services, social functioning, recreation or leisure activity, sleep functioning, nonuse of health care services, and hemodynamic 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 (i.e., pathophysiologic 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, the 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.

Most of the domains we identified (75%) were measured at similar frequencies across all of the included shoulder disorders (including pain, physical function, and range of motion). Because of this, we think it would be appropriate to develop a single core outcome set for all shoulder disorders, rather than developing different sets for different disorders (e.g., one for rotator cuff disease and a different one for adhesive capsulitis). A single core outcome set would not preclude trialists from measuring other domains that might be especially relevant in the context of the particular trial (13,27). For example, a trial of surgical repair of rotator cuff tears may use imaging to assess the integrity of the repair, whereas 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 the time to return to sport. However, we recognize that the decision to create a single core outcome set for all shoulder disorders requires consideration from various stakeholders, the most important of which are the patients, who can help ensure that the core outcome set sufficiently captures the experience of their condition (28).

Government sponsorship of initiatives such as the Patient-Centered Outcomes Research Institute indicates that

policymakers 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, which was used in 31% of the trials, includes physician-rated components to measure range of motion and strength (impairment measures). It is unclear whether trialists believe that pain is the only PROM that should be measured 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 the 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 core outcome set for shoulder disorders.

Additional research is needed before we can recommend a core outcome set 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 changes in symptoms/function over time. A systematic review of the measurement properties of instruments for shoulder disorders will help to determine which are most fit for this 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, and the endorsement of OMERACT will be sought.

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be submitted for publication. Dr. Buchbinder had full access to all of the data in the study and

takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study conception and design. Page, Huang, Gagnier, Buchbinder.

Acquisition of data. Page, Huang.

Analysis and interpretation of data. Page, Huang, Verhagen, Gagnier, Buchbinder.

REFERENCES

- 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 (Oxford)* 2006;45:215–21.
- 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;44:800–5.
- Luime JJ, Koes BW, Hendriksen IJ, Burdorf 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.
- 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.
- 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 Musculoskelet Disord* 2012;13:238.
- Van der Heijden GJ. Shoulder disorders: a state-of-the-art review. *Baillieres Best Pract Res Clin Rheumatol* 1999;13:287–309.
- 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 2,674 patients. *J Shoulder Elbow Surg* 2006;15:30–9.
- Mather RC III, 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;95:1993–2000.
- 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.
- Clarke M. Standardising outcomes for clinical trials and systematic reviews. *Trials* 2007;8:39.
- 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.
- 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 [editorial]. *J Health Serv Res Policy* 2012;17:1–2.
- Boers M, Kirwan JR, Wells G, Beaton D, Gossec L, d'Agostino MA, et al. Developing core outcome measurement sets for clinical trials: OMERACT filter 2.0. *J Clin Epidemiol* 2014;67:745–53.
- 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:e0146444.
- 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. *J Rheumatol* 1994;S41:86–9.
- 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;1–16.
- Schumacher HR Jr, Edwards LN, Perez-Ruiz F, Becker M, Chen LX, Furst DE, et al. Outcome measures for acute and chronic gout. *J Rheumatol* 2005;32:2452–5.
- 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;68:1270–81.
- Gagnier JJ, Page MJ, Huang H, Verhagen AP, Buchbinder R. Creation of a core outcome set for clinical trials of people with shoulder pain: a study protocol. *Trials* 2017;18:336.
- Whittle S, Buchbinder R. In the clinic: rotator cuff disease. *Ann Intern Med* 2015;162:1–15.
- Schellingerhout JM, Verhagen AP, Thomas S, Koes BW. Lack of uniformity in diagnostic labeling of shoulder pain: time for a different approach. *Manual Ther* 2008;13:478–83.
- World Health Organization. International Classification of Functioning, Disability and Health (ICF) index. Geneva: World Health Organization; 2001.
- StataCorp. Stata Statistical Software: release 13. College Station (TX): StataCorp; 2013.
- Constant CR, Murley AH. A clinical method of functional assessment of the shoulder. *Clin Orthop Relat Res* 1987; 214:160–4.
- Roach KE, Budiman-Mak E, Songsiridej N, Lertratanakul Y. Development of a shoulder pain and disability index. *Arthritis Care Res* 1991;4:143–9.
- Dwan K, Gamble C, Williamson PR, Kirkham JJ, and 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.
- 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.
- 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. *J Rheumatol* 2005;32:2250–6.
- 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. *J Comp Eff Res* 2016;5:193–205.
- 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;96:265–71.
- 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;98:251–6.
- 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 asymptomatic atraumatic full-thickness rotator cuff tear. *J Bone Joint Surg Am* 2014;96:793–800.
- 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. *Orthop J Sports Med* 2015;3:2325967115584596.
- Black N. Patient reported outcome measures could help transform healthcare. *BMJ* 2013;346:f167.
- 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; 17:230.