

## RESEARCH ARTICLE

## Open Access



# A narrative review and content analysis of functional and quality of life measures used to evaluate the outcome after TSA: an ICF linking application

Ze Lu<sup>1,2\*</sup>, Joy C. MacDermid<sup>1,2,3</sup> and Peter Rosenbaum<sup>4</sup>

## Abstract

**Background:** Total shoulder arthroplasty (TSA) is considered as the standard reconstructive surgery for patients suffering from severe shoulder pain and dysfunction caused by arthrosis. Multiple patient-reported outcome measures (PROMs) have been developed and validated that can be used to evaluate TSA outcomes. When selecting an outcome measure both content and psychometric properties must be considered. Most research to date has focused on psychometric properties. Therefore, the current study aims to summarize what PROMs are being used to assess TSA outcomes, to classify the type of measure (International society for quality of life (ISOQOL) using definitions of functioning, disability, and health (FDH), quality of life (QoL) and health-related quality of life (HRQoL)) and to compare the content of these measures by linking them to the International Classification of Functioning, Disability and Health (ICF) framework.

**Methods:** A literature review was performed in three databases including MEDLINE, EMBASE, and CINAHL to identify PROMs that were used in TSA studies. Meaningful concepts of the identified measures were extracted and linked to the relevant second-level ICF codes using standard linking rules. Outcome measures were classified as being FDH, HRQoL or QoL measures based on the content analysis.

**Result:** Thirty-five measures were identified across 400 retrieved studies. The most frequently used PROM was the American Shoulder and Elbow Society score accounting for 21% (246) of the total citations, followed by the single item pain-related scale like visual analog scale (17%) and Simple Shoulder Test (12%). Twelve PROMs with 190 individual items fit inclusion criteria for conceptual analysis. Most codes (65%) fell under activity and participation categories. The top 3 most predominant codes were: sensation of pain (b280; 13%), hand and arm use (d445; 13%), recreational activity (d920; 8%). Ten PROMs included in this study were categorized as FDH measures, one as HRQoL measure, and one as unknown.

(Continued on next page)

\* Correspondence: [luz35@mcmaster.ca](mailto:luz35@mcmaster.ca)

<sup>1</sup>Roth|McFarlane Hand and Upper Limb Centre, St Joseph's Health Care, London, ON, Canada

<sup>2</sup>The School of Rehabilitation Science, McMaster University, Hamilton, ON, Canada

Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

(Continued from previous page)

**Conclusions:** Our study demonstrated that there is an inconsistency and lack of clarity in conceptual frameworks of identified PROMs. Despite this, common core constructs are evaluated. Decision-making about individual studies or core sets for outcome measurement for TSA would be advanced by considering our results, patient priorities and measurement properties.

**Keywords:** Patient-reported outcome measures, Total shoulder arthroplasty, ICF, Health-related quality of life, Quality of life

## Background

The high prevalence of shoulder pain (7–21%) in the general population results in high, and increasing, levels of disability and health-care costs [1, 2]. Glenohumeral arthritis is the primary cause of shoulder pain and dysfunction in an aging population [1]. Psychological issues such as depression, anxiety, and decreased quality of life (QoL) are associated with chronic musculoskeletal pain [1–4]. The combined physical and psychosocial impacts of shoulder pain are complex and contribute to lower quality of life [2].

Total shoulder arthroplasty (TSA) is reconstructive surgeries that can provide pain relief and restore function in severely damaged arthritic shoulders [5–8]. Such treatments have a predictable outcome for patients with joint destruction arising from pathologies such as osteoarthritis, rheumatoid arthritis and proximal humeral head fracture [8–10]. Previous studies indicated significant improvement of both psychological status and health-related quality of life (HRQoL) by 3-months after surgery [2, 11]. However, implant issues, such as loosened glenoid components can lead to poorer outcomes over the longer-term [3, 7]. While improvement can be expected, normal function cannot be restored and the outcome achieved is variable and dependent on many factors including different surgical indications, soft tissue recovery, subscapularis integrity, and post-operative rehabilitation [5, 12].

To evaluate surgery outcomes, many clinicians and researchers are aware of the importance of measuring pain, functional outcomes, biopsychological health, QoL, and HRQoL [13]. Since the 1990s, numerous patient-reported outcome measures (PROMs) have been developed and validated to assess outcomes in shoulder conditions [14–16]. Researchers have established acceptable levels of reliability, validity, and responsiveness of PROMs such as the American Shoulder and Elbow Society score (ASES) and the Simple Shoulder Test (SST), and synthesized the evidence of psychometric properties in a systematic way [7, 13, 17, 18].

Content validity is a fundamental property of PROM, but relatively unintended to in the literature. Although standard definitions exist for functioning, disability, and health (FDH), rather than HRQoL or QoL [19–21], there is overlap in these concepts and insufficient precision by

developers and users with respect to these terms [20]. The absence of a theoretical framework or conceptual definition leads to difficulty of interpreting study results using different outcome measures, since it can be unclear which domains of health are affected by the intervention, and whether differences in outcomes relate to the intervention or the measure [5]. Since few developers provide clear definitions of their latent constructs or how items were mapped to these constructs it is important to do a retrospective evaluation to inform content validation and understand differences in constructs evaluated by commonly used measures.

The World Health Organization (WHO) definitions provided in the International Classification of Functioning, Disability and Health (ICF) and the manual of WHO-Quality of Life brief version (WHOQOL-BREF) [21–24] provided internationally used frameworks, definitions and coding language to describe the impact of health conditions on function, disability and health [21]. With the ICF framework, the concept of FDH refers to the biopsychosocial components and interactions among body structures and function, and activities and participation in the context of the environment and personal factors [25–28]. The QoL is defined by WHO as “a person’s perception of their position in life affected by the culture and value system in which they live and in relation to goals, expectations, standards, and concerns [21, 23].”

However, the terminology HRQoL, remains variably defined by different sources [20]. In HRQoL is specific focus on how QoL is influenced by a health condition [20, 23]. While culture, politics and economic context also affect QoL, those influences are generally not addressed in HRQoL measures or health-related PROM [20]. See Table 1 for organizations of the concepts for FDH, QoL, and HRQoL.

Using the ICF framework, researchers can evaluate individual items of content, by a standardized coding, or “linking” procedure [29, 30]. According to the ICF linking rules, a second level codes start with the letters b, s, d and representing the classification of body function and structure, activity, participation, environmental factors and personal factors followed by a numeric code for the chapter number (one digit) and another two digits as the second level [29]. The linking process is a universal

**Table 1** The organizations of concepts of functioning, disability, and health (FDH), quality of life (QoL), and health-related quality of life (HRQoL)

Concept	Definition adopted	Visual explanation	Item example (name of measurement)
Functioning, disability, and health (FDH)	Biopsychosocial components and interactions among body structures and function, and activities and participation in the context of the environment and personal factors	<p>The diagram illustrates the International Classification of Functioning, Disability and Health (ICF) model. At the top is 'Health Condition (disorder or disease)'. Below it are three interconnected components: 'Body Functions &amp; Structure', 'Activity', and 'Participation'. 'Activity' is centrally located and has bidirectional arrows connecting it to 'Body Functions &amp; Structure' and 'Participation'. Below these components are 'Environmental Factors' and 'Personal Factors', both of which have arrows pointing upwards to influence the 'Activity' component.</p>	Is it difficult for you manage toileting? (American Shoulder and Elbow Society Score)
Quality of life(QoL)	A person's perception of their positions in life affected by the culture and value system in which they live and in relation to goals, expectations, standards, and concerns	<p>This diagram shows 'Health, Economy, Culture, Policy, Value, Resources, etc.' in a central box. It is surrounded by four ovals: 'Standards' (top-left), 'Concerns' (top-right), 'Expectations' (bottom-left), and 'Goals' (bottom-right). Arrows point from each of these ovals towards the central box, indicating their influence on the overall health and quality of life.</p>	How much of a burden do you feel you are on others? (Western Ontario Osteoarthritis Score)
Health-related quality of life (HRQoL)	The subjective assessment of the impact of disease and treatment across the physical, psychological, social and somatic domains of functioning and well-being	<p>This diagram shows 'Functioning, disability, and health (FDH)' in a central box. It is surrounded by four ovals: 'Standards' (top-left), 'Concerns' (top-right), 'Expectations' (bottom-left), and 'Goals' (bottom-right). Arrows point from each of these ovals towards the central box, indicating their influence on the functional and health status.</p>	How satisfied are you with the current level of function of your shoulder? (PENN shoulder score)

language that can be used to define the content of items and describe their meaningful constructs [26]. A systematic review conducted in 2013 evaluated the content of 475 shoulder related outcome measures by linking individual items to ICF codes [18]. This work provided some evidence on content validity of shoulder PROM. We are building on this work by focusing on shoulder arthroplasty to understand the use of PROM in this area of practice, providing an updated assessment of PROM usage, and classifying the conceptual framework according to standard definitions.

The objective of the current study was to analyze the classification and content of functional and quality of life measures used to evaluate the outcome after TSA using the ICF framework by (1) identifying the PROMs used for patients after TSA; (2) mapping the content of the individual items using second level ICF codes; (3) summarizing the focus of these PROMs based on ICF domains; and (4) providing an updated assessment of

PROM usage and summarizing the predominant application of included PROMs based on ICF linking and pre-defined concepts of FDH, HRQoL, and QoL.

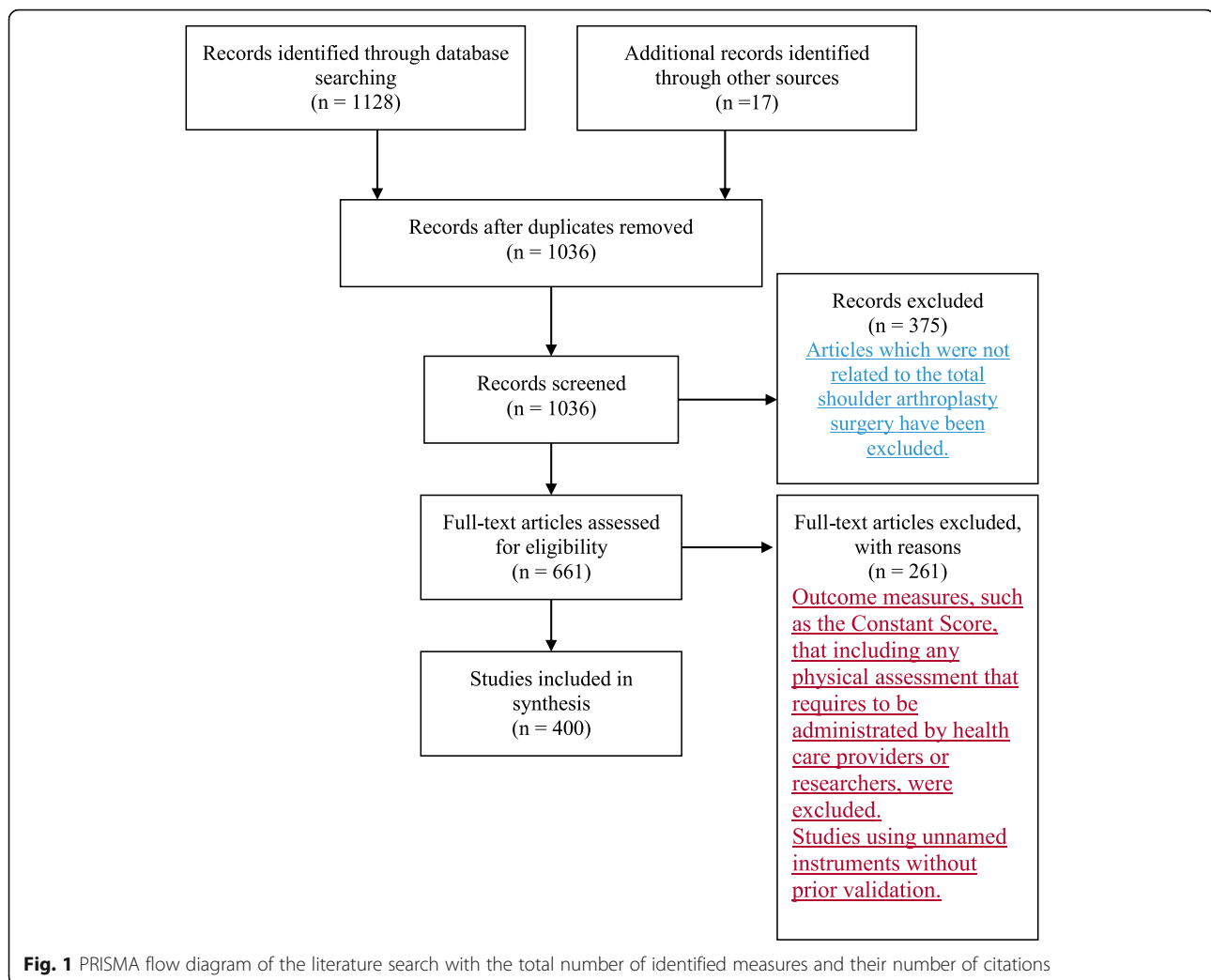
**Methods**

**Design**

A structured literature review was carried out following the PRISMA guideline [31]. The PRISMA flow diagram containing all steps of the screening and extraction of measures are displayed in Fig. 1. The content analysis of PROMs used for patients post TSA surgery was performed based on the existing ICF linking rules [29, 30].

**Information sources**

A literature search was performed in three databases including MEDLINE, EMBASE, and CINAHL to capture PROMs used for patients with TSA in both clinical and research settings.



## Search

MeSH terms were used for PROMs, including questionnaire, score, index, tool, survey, outcome measure, and patient-report, were connected by Boolean operator 'OR.' The same operator was applied for other MeSH terms that specified TSA management by total shoulder arthroplasty and total shoulder replacement. PROMs and TSA terms were then combined with the operator 'AND' for final search. We limited searching to the last 5 years and 3 months, from January 2014 to December 2019, to reflect recent practice. Details of search keywords are listed in Additional file 1.

## Eligibility criteria

The inclusion criteria were any articles published in peer-reviewed journals on studies of TSA surgery using named PROMs to measure FDH, QoL or HRQoL. Outcome measures, such as the Constant Score, that including any physical assessment that requires to be administrated by health care providers or researchers, were excluded since

our focus was PROM. We also excluded studies using unnamed instruments without prior validation.

## Study selection

All identified articles were imported into Mendeley reference management software (version 1.19., 2008 Glyph & Cog, LLC) for duplicate, author and journal information checking. After removal of the duplication, the first author [ZL] performed the title, abstract and full-text review. At full-text review stage, the second author [JMacD] randomly reviewed 50% of the articles and discussed the disagreement with the first author through regular meetings.

## Data collection process

Data extraction was initially performed by the first author [ZL]. The original intention of using the instrument (e.g., to measure pain, function, QoL, HRQoL, surgery outcome, patient satisfaction, etc.) was recorded. Ambiguous or difficult cases were presented through online-based discussion for the final decision. We calibrated the details of

PROMs, such as different versions of the questionnaires, according to a previous systematic review and a guideline of shoulder outcomes measures [18]. To avoid detailed analysis of rarely used PROM, we excluded PROMS that had not been cited at least 10 times from the overall data pool of 1175 citations thereby excluding those that represent less than 1%. The tracking sheet of excluded questionnaires is available upon request.

#### Data items

According to the predetermined definitions of FDH, QoL, and HRQoL, the original intention of applying the PROMs was recorded and analyzed through the data extraction [23]. The first author documented text in the articles where they referred to researchers' purpose of using PROMs and then coded the outcome measures for different conceptual applications. Direct clarifications in terms of function, disability or health, QoL or HRQoL were categorized onto terms FDH, QoL, or HRQoL. Ambiguous statements were coded with the consideration of the context of studies. For example, patients' satisfaction level with their shoulder condition was coded as HRQoL.

#### Summary measures (content analysis)

The content of included PROMs was evaluated item by item based on existing ICF linking rules [23, 29, 30]. One of the authors [ZL] finished the entire linking work independently and then presented the result to an external expert with experience in ICF. Any discrepancies were marked as addressed if agreement was achieved. Meaningful concepts were linked to the specific second level of the ICF codes. An individual item can map onto several codes if needed. For example, *pain pushing with the involved arm* contains meaningful concepts as *pain* and *pushing with involved arm*, which was coded separately as *sensation of pain* (b280) and *hand and arm use* (d445). General concepts that cannot be assigned with a code but are still within the classification system were linked as *non-definable* [18, 23, 29]. For example, the general evaluation of the health condition was coded as *nd* due to the coverage of all aspects of health without specific definitions. *Not covered* (*nc*) was used for the concepts beyond the ICF conceptual framework, such as the satisfaction level about the quality of health care Personal factor was labeled as *pf*, and consistent with ICF were acknowledged, but not coded.

We used summary indices to decide the extent to which content of a measure can be captured with ICF codes [32, 33]. The formula was listed as follows: The number of items linked to at least one ICF code/total number of items on the measure  $\times 100\%$ .

#### Synthesis of results (from the content analysis)

Individual item codes were then categorized into the five ICF domains, including *body function and structure*, *activity, participation*, *environmental factors* and *personal factors* according to the linking result.

As the final step, included PROMs were summarized into FDH, HRQoL and QoL measures with the recommended use based on the previous analysis. Measures focusing on pain, shoulder function, capacity, performance, difficulty, barriers or facilitators of contextual factors were categorized as FDH measures. The dominant perspectives of measures were provided based on content analysis using ICF. Other questionnaires that mainly ask expectations, evaluation, and person judgment about health or health-related domain were coded as HRQoL. QoL measures were also classified based on the WHO definition. FDH measures with HRQoL/QoL features were given when at least one item from FDH scales was not covered by ICF component but within the HRQoL/QoL. For example, if one item from a given measure was categorized onto HRQoL related content, while other questions were all identified as FDH, this specific PROM was considered as a FDH measure with HRQoL features.

Previous evidence from the literature review was cross-referenced at this stage. Consensus was required from all three authors to finalize the result [25, 33].

## Results

#### Study selection

Overall, 1036 studies were screened through the title and abstract review, and 400 of these articles were included. We identified thirty-five measures that have been cited 1175 times from all retrieved studies. Among them, five were single item questionnaires, and 30 were multi-item measures. Please see Additional file 2 for all 35 outcome measures. The Constant and three other non-PROMs, including the Constant-Murley and Charlson Morbidity Index, were not involved in further content analysis. Numeric rating scales (NRS) and visual analog scales (VAS) for pain were considered as the same measure due to similar meaning of the content of the question. The same strategy was applied for the Single Assessment Numeric Evaluation (SANE) and Subjective Shoulder Value (SSV). All studies used an English version of the PROMs. In total, 12 PROMs for our inclusion and exclusion criteria and underwent detailed ICF linking and conceptual analysis.

#### Results of individual studies (second level of ICF linking)

A total of 36 s level ICF codes were linked to individual items (Table 2). There were 23 different codes under the *activities and participation* category (*d* codes) and 10 under the *body structure* (*s* codes) and *body function* (*b* codes). *Personal factors* were identified within three

**Table 2** Second level ICF categories linked to the individual items from included PROMs in ranked order

ICF code linked to items Identified PROMs (Number of items)	b280 Sensation of pain	d445 Hand and arm use	d920 Recreation and leisure	d850 Remunerative employment	d430 Lifting and carrying objects	d640 Doing household work	b730 Muscle power functions	d540 Dressing	d510 Washing oneself	d230 Carrying out daily routine	b134 Sleep functions
ASES [17]	4	2	2	2	1		1	1	1		1
VAS&NRS-pain	1										
SST [12]	1	5		1	3		1	1	1		1
SANE & SSV [1]	<b>None</b>										
SF-12 (12)	1		2	4		2				4	
SPADI [13]	5	3			2		1	3	2		
DASH [34]	4	7	7	5	1	5	2	1	2	1	1
WOOS [19]	4	3	1		1	1	1	2		1	1
Quick-DASH [19]	4	2	5	5	1	4	1		1	1	1
OSS [12]	4	1		1	1	2		1	1		
PSS [26]	3	5	2	1	7	1	4	3	3		1
Patient satisfaction [1]	<b>Not covered</b>										

In a total, 36 s level ICF codes were used to link individual items. Eleven of the total linked codes were found with a frequency above 5%.

- ASES American Shoulder and Elbow Society
- SST Simple Shoulder Test
- DASH Disabilities of the Arm, Shoulder, and Hand
- WOOS Western Ontario Osteoarthritis Score
- PSS PENN shoulder score
- NRS Numerous Numeric rating scales
- VAS Visual analog pain scales
- SANE Single Assessment Numeric Evaluation
- SSV Subjective Shoulder Value
- SPADI Shoulder Pain and Disability Index
- OSS Oxford Shoulder Scale



included PROMs: The Disabilities of the Arm, Shoulder, and Hand (DASH), Western Ontario Osteoarthritis Score (WOOS), and PENN shoulder score (PSS). Only two codes under environment factors were identified as *products or substances for personal consumption* (e110) and *climate* (e225). Eleven of the total linked codes were found with a frequency above 5% as *sensation of pain* (b280), *hand and arm use* (d445), *recreation and leisure* (d920), *remunerative employment* (d850), *lifting and carrying objects* (d430), *doing housework* (d640), *muscle power functions* (b730), *dressing* (d540), *washing oneself* (d510), *carrying out daily routine* (d230), and *sleep functions* (b134). The occasions of using these codes to link individual item in each PROMs were listed in rank order in Table 2.

Of all the measures, one item proposed as “Since beginning therapy for your shoulder, would you say that your shoulder has” from PSS could not be linked by specific categories but was considered still within the ICF framework (nd). Six PROMs including SANE, SSV, SF-12, DASH, WOOS, and PSS had a question that was not covered by the ICF but was within HRQoL. One item of WOOS, asking how much of a burden do you feel you are on others, was categorized as QoL-related content. A summary of the distribution of items from each PROM under the ICF chapter level is listed by frequency order in Table 3.

**Synthesis of results (summarization of predominant application)**

An overview of the summarized information for each PROMs is presented in Table 4. The most frequently used PROM was the American Shoulder and Elbow Society score (ASES) accounting for 21% (246 times) of the total citations, followed by the NRS or VAS for 17% and SST for 12%. Most of the analyzed measures were used as functional outcome instruments. Through the review, we found that the SF-12 was often used as a tool to evaluate QoL, although it was designed as a health status measure. Patient satisfaction scales were used to quantify the personal expectation to the surgery, care or shoulder conditions.

The high percentage of the measure to ICF linkage indicated that most of the items from included PROMs can be linked with second-level ICF, except for SANE/SSV and patient satisfaction, which are not linkable constructs. Ten of the PROMs included in this study were categorized as FDH measures, with specific focus of quantifying symptoms and functional limitations for people with shoulder problems.

**Discussion**

This study found variation between commonly used PROM used to assess the outcomes of TSA in terms of their overall latent construct and the item level content, although most were more focused on activity and

**Table 3** Categorization of items under ICF domains with corresponding percentage

	Body function and structure (25%)	Activity (41%)	Participation (24%)	Environment (2%)	Personal factors (2%)	Not defined but within ICF (1%)	QOL/HRQOL (5%)
ASES	2	8	4	3			
VAS&NRSpain	1						
SST	5	9	1				
SANE & SSV							1
Patient satisfaction							
SF-12	3	1	7				3
SPADI	6	12					
DASH	10	17	16		1		1
WOOS	7	7	2	1	2		3
Quick-DASH	6	4	12				
OSS	4	7	2				
PSS	7	18	5		1	1	2

ASES American Shoulder and Elbow Society  
 SST Simple Shoulder Test  
 DASH Disabilities of the Arm, Shoulder, and Hand  
 WOOS Western Ontario Osteoarthritis Score  
 PSS PENN shoulder score  
 NRS Numerous Numeric rating scales  
 VAS Visual analog pain scales  
 SANE Single Assessment Numeric Evaluation  
 SSV Subjective Shoulder Value  
 SPADI Shoulder Pain and Disability Index  
 OSS Oxford Shoulder Scale

**Table 4** Description of the FDH, HRQoL, and QoL perspective based on ICF linking

	Percentage of total citations	Measure to ICF linkage	Dominant ICF Component	Dominant intention	Instrument recommendations
<b>ASES</b>	21%	100%	Activity	Functional outcome	FDH instrument focusing on activity concerns
<b>VAS-pain</b>	17%	100%	Body function	Pain evaluation	FDH instrument for pain
<b>SST</b>	12%	100%	Activity Body function	Functional outcome	FDH instrument focusing on activity and body function
<b>SANE &amp; SSV</b>	8%	0	none	Functional outcome	HRQoL
<b>Patient satisfaction</b>	5%	0	none	Patient satisfaction	Unknown
<b>SF-12</b>	3%	92%	Participation	QoL	FDH instrument focusing participation with HRQoL feature
<b>SPADI</b>	3%	100%	Activity Body function	Functional outcome	FDH instrument focusing on activity and body function
<b>DASH</b>	3%	100%	Activity Participation Body function	Functional outcome	FDH instrument with HRQoL
<b>WOOS</b>	2%	95%	Activity Body function	Functional outcome	FDH instrument focusing on activity and body function with HRQoL and QoL feature
<b>Quick-DASH</b>	2%	100%	Participation	Functional outcome	FDH instrument focusing on participation
<b>OSS</b>	2%	100%	Activity	Functional outcome	FDH instrument focusing on activity concerns
<b>PENN</b>	1%	92%	Activity	Functional outcome	FDH instrument focusing on activity with HRQoL feature

FDH Functioning, disability and health

HRQoL Health-related quality of life

QoL Quality of life

ASES American Shoulder and Elbow Society

SST Simple Shoulder Test

DASH Disabilities of the Arm, Shoulder, and Hand

WOOS Western Ontario Osteoarthritis Score

PSS PENN shoulder score

NRS Numerous Numeric rating scales

VAS Visual analog pain scales

SANE Single Assessment Numeric Evaluation

SSV Subjective Shoulder Value

SPADI Shoulder Pain and Disability Index

OSS Oxford Shoulder Scale

participation than patients perceptions of body structure and function. Overall, the content covered by the PROM included 10 s level ICF codes under the domain of *body functions and structures*, and 23 codes belong to *activities and participation*. This is consistent with the fact that PROMs focus on the patient be experience and uniquely able to assess how a person functions in their own life; whereas impairments in body structure and function can be better measured with clinical tests. Only two categories under *Environmental factors* were mentioned. Other content analysis of PROMs has noted a similar lack of attention to the environment [18]. Even where environment is not explicitly addressed, we expect it to be an important factor in disability that may partially explain why patients with similar impairments experiences different disability.

Pain is a primary concern for patients with TSA surgery [6]. This is consistent with that one category the *sensation of pain (b280)*, was ranked as the most frequently used code. Although pain is considered an impairment in ICF, it also a subjective experience and as

such typically captured by PROM. The second most linked code under *body function* domain was *muscle power function (b730)*, which belongs to the impairment domain under the ICF framework. Strength can be assessed by clinicians using dynamometers or other devices; or can be self-reported by patients. Typically, we expect that PROM would focus on functional items and that pain, motion and strength might all interfere with functional performance. However, some PROMs do ask questions that specifically target muscle strength. Generally, these questions must be fairly generic rather than target specific muscle groups as might be assessed by dynamometers For example, questions from SST that ask participants to rate the difficulty of lifting task with three pre-defined weight levels ranging from one lb. to 20 lbs., assess strength, but do not identify particular muscle groups or adaptations.

*Activity* was the predominant ICF domain, accounting for 41% of the items [21, 28]. *Hand and arm use and lifting and carrying objects* are the most commonly linked ICF codes under the *activity* domain. This



suggests a consistent recognition of the importance of these tasks in patients with shoulder arthritis, requiring TSA. However, on the other hand, the ICF categories related to mental function such as *sleep function, emotional function, and energy and drive* were infrequently linked suggesting less agreement that these are central to TSA outcomes. Given the importance of psychological health [18, 35] in post-surgical patients, one might consider this as under-representation, especially if the instrument is intended to measure QoL. However, outcome instrument developers often try to focus on a clear construct, and it would be the responsibility of researchers to include measures of physical health and psychologic health within their studies, since summing different construct together may not always be appropriate. Further, developers may consider psychological factors as mediators of outcomes rather than the outcomes themselves. Ideally, developers would be explicitly explaining these conceptual assumptions.

*Recreation and leisure* (d920) and *Remunerative employment* (d850) were ranked as third and fourth order among all the linked items. This high ranking is consistent with a previous systematic review focusing PROMs of shoulder pain and functioning [18]. Most PROMs such as ASES, DASH, and Oxford Shoulder Scale (OSS) imply these concepts by formulating questions as leisure activities and usual work. Overall, these items and others that fit within participation comprise 24% of the total items. The concept of *participation* defined by the ICF framework is subject to qualifiers that describe what a person does in their usual life. That means subjects' response to such questions might be modified by the usual roles or environmental factors, but these are not directly measured.

According to the WHO, different PROMs used for patients after TSA share areas of content and purposes of application. The single item measure SANE, and SSV, that investigate to what extent a patient would rate their shoulder as being normal, was classified as HRQoL perspective since it address a global evaluation, whereas researchers and clinicians commonly use it as a functional outcome since it assumed to be rating physical function on a scale of 0–100% [36]. A previous study found that patients have a lot of confusion about what is being calibrated when responding to this questions, which reflect the ambiguity in its definition [37]. It is important to have a conceptual distinction between measures designed to assess HRQoL which is intended to be comprehensive, versus those designed to measure physical functioning which a smaller construct that might affect QoL. The confusion we found in conceptual clarity and content of items across many of these measures emphasizes the importance for instrument developers to define their conceptual framework so that users of outcome

measures can match the measurement purpose is to a specific conceptual framework.

Patients satisfaction is important, but often variably measured in health research. Satisfaction with care is a process measure; whereas as satisfaction with health/shoulder status can be considered as an HRQoL measure. However, by asking about the satisfaction with surgery and care, this scale mixes evaluation of the process of care or Quality of Care [34], with outcome evaluation. Researchers and clinicians should be more explicit about whether they are measuring process or outcome satisfaction; and ensuring their selected measure reflects that choice.

A key issue in the literature was the vague and imprecise terminology used to for different outcome measures and the definition of the FDH, QoL, and HRQoL. For researchers, clearly defined concepts within PROMs help them detect the most appropriate and precise latent construct. For clinicians, in both research and daily practice work, appropriate selection of the outcomes measures is not only depend on the psychometric properties and intention of the application, but also on the precise understanding of the content informed by an unified conceptual framework [25]. Developers rarely provide a strong conceptual framework, and users rarely state their measurement rationale or the content validity of the tools they selected for the constructs of interest. Rather justification of PROMs within studies tend to focus on psychometric properties like reliability, which do not reflect content validity. Some measures mix different constructs. For example, The DASH provides a comprehensive set of items and is defined an as FDH instrument but contains items that fall within a HRQoL construct. Terms are often used incorrectly, for example health status measures and functional measures are often referred to as QoL measures. The use of terms like clinical outcome measures or functional outcomes happens without clear distinction about what these terms mean [11, 38]. The conceptual analysis performed in the current study may help resolve the issue by precisely categorizing the retrieved PROMs into three types as: (1) FDH (the capacity, performance, presence / absence, frequency, severity, or other biopsychosocial domains), (2) HRQoL (the expectations, standards, or concerns about individual health), and (3) QoL (the patient's personal assessment of their position in life). Mapping the ICF domains within PROM can help researchers and clinicians to select the most appropriate PROMs for their context (and research question). That is considering the impacts of shoulder joint destruction (or indications for TSA) and expected impacts of TSA (outcomes) should drive the PROM that have the best conceptual match. Further, this can identify when important constructs are missing, and

supplemental measures might be needed. This would complement, not replace, considering important psychometric properties like reliability and responsiveness.

TSA outcomes measures should be developed under a clear conceptual framework. Many of the consensus panels that attempt to achieve consensus on outcome measurement start with defining the core constructs that should be measured for a given health problem, and then choose the best measure within those constructs [25, 33, 39]. Our findings could support such a process. A better understanding of the latent construct evaluated within PROM is essential to enable clinicians and researchers to make valid conclusions. For those validated PROMs, clinician should also be cautious to use them in different conditions such as other language versions. The cross-cultural adaptation might not be able to ensure the content validity with the consideration of the various culture background, healthcare systems.

### Limitation

The current review does have limitations. Our search strategy may not have identified all studies using PROMs. However, the large number of studies we reviewed created robust findings. Our exclusion of rarely used PROM may have missed some emerging but higher quality PROM that have different or more clear content validity. Extraction of data was complicated by a lack of clear reporting in some papers. Even with the updated version of ICF linking rules, personal factors are still not classified F [29], and so while we acknowledge these as important they were not classified. ICF coding is one approach to assess content validity and should be supplemented by other methods including cognitive interviews and quantitative patient/expert ratings of relevance.

### Conclusion

We found confusion in conceptual definitions on PROMs, and wide variation in PROM content and use. Despite the variability there were some common constructs evident in measurement of pain, hand and arm use, recreational activities work and employment, lifting and carrying. Mental function components such as emotional function, and energy and drive were rarely covered reflecting the focus on physical recovery following TSA. Users evaluated in these constructs may require supplemental PROM. Efforts to the consensus on the key constructs that should be measured following TSA are needed.

### Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s12891-020-03238-w>.

**Additional file 1.**

**Additional file 2.** The list of all 35 outcome measures.

### Abbreviations

PROMs: Patient-reported outcome measures; ICF: International Classification of Functioning, Disability and Health; ISOQOL: International society for quality of life; FDH: Functioning, disability, and health; WHO: World Health Organization; WHOQOL-BREF: WHO-Quality of Life brief version; HRQoL: Health-Related Quality of Life; QoL: Quality of Life; TSA: Total shoulder arthroplasty; ASES: American Shoulder and Elbow Society; SST: Simple Shoulder Test; DASH: Disabilities of the Arm, Shoulder, and Hand; WOOS: Western Ontario Osteoarthritis Score; PSS: PENN shoulder score; NRS: Numeric rating scales; VAS: Visual analog scales; SANE: Single Assessment Numeric Evaluation; SSV: Subjective Shoulder Value; SPADI: Shoulder Pain and Disability Index; OSS: Oxford Shoulder Scale

### Acknowledgements

Joy MacDermid was supported by a Canadian Institutes of Health Research Chair in Gender, Work and Health and the Dr. James Roth Chair in Musculoskeletal Measurement and Knowledge Translation. The authors would like to thank Dr. Olaf Kraus de Camargo, Dr. Jan Willem Gorter, and other members at the *CanChild* Centre for Childhood Disability Research at McMaster University for their support and valuable feedback during the conduct of the study.

### Authors' contributions

ZL completed the literature review, content analysis, data interpretation, and was a major contributor in drafting the manuscript. PR help finalized the topic of the study and data interpretation. JMacD contributed to the current study as one of the reviewers during literature search, ICF linking. PR and JMacD both provided feedback for writing manuscript. All authors read and approved the final manuscript.

### Funding

Not applicable.

### Availability of data and materials

Not applicable.

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

Not applicable.

### Author details

<sup>1</sup>Roth|McFarlane Hand and Upper Limb Centre, St Joseph's Health Care, London, ON, Canada. <sup>2</sup>The School of Rehabilitation Science, McMaster University, Hamilton, ON, Canada. <sup>3</sup>Physical Therapy and Surgery, Western University, London, ON, Canada. <sup>4</sup>CanChild Centre for Childhood Disability Research, McMaster University, Hamilton, ON, Canada.

Received: 4 May 2019 Accepted: 25 March 2020

Published online: 13 April 2020

### References

1. Cho CH, Jung SW, Park JY, Song KS, Yu KI. Is shoulder pain for three months or longer correlated with depression, anxiety, and sleep disturbance? *J Shoulder Elb Surg*. 2013;22(2):222–8. <https://doi.org/10.1016/j.jse.2012.04.001>.
2. Cho CH, Song KS, Hwang I, Coats-Thomas MS, Warner JJP. Changes in psychological status and health-related quality of life following total shoulder arthroplasty. *J Bone Jt Surg - Am Vol*. 2017;99(12):1030–5. <https://doi.org/10.1016/j.jse.2016.12.018>.
3. Carter MJ, Mikuls TR, Nayak S, Fehring EV, Michaud K. Impact of total shoulder arthroplasty on generic and shoulder-specific health-related quality-of-life measures: a systematic literature review and meta-analysis. *J Bone Jt Surg - Ser A*. 2012;94(17):1–9.
4. Henn RF, Ghomrawi H, Rutledge JR, Mazumdar M, Mancuso CA, Marx RG. Preoperative patient expectations of total shoulder arthroplasty. *J Bone Jt Surg - Ser A*. 2011;93(22):2110–5.

5. Roy J-S, Macdermid JC, Goel D, Faber KJ, Athwal GS, Drosdowech DS. What is a successful outcome following reverse Total shoulder Arthroplasty? *Open Orthop J*. 2010;4:157–63.
6. Radnay CS, Setter KJ, Chambers L, Levine WN, Bigliani LU, Ahmad CS. Total shoulder replacement compared with humeral head replacement for the treatment of primary glenohumeral osteoarthritis: a systematic review. *J Shoulder Elb Surg*. 2007;16(4):396–402.
7. Bryant D, Litchfield R, Sandow M, Gartsman GM, Guyatt G, Kirkley A. A comparison of pain, strength, range of motion, and functional outcomes after hemiarthroplasty and total shoulder arthroplasty in patients with osteoarthritis of the shoulder: A systematic review and meta-analysis. *J Bone Jt Surg - Ser A*. 2005;87(9 I):1947–56.
8. Heuberger PR, Brandl G, Pauzenberger L, Laky B, Krieglleder B, Anderl W. Radiological changes do not influence clinical mid-term outcome in stemless humeral head replacements with hollow screw fixation: a prospective radiological and clinical evaluation. *BMC Musculoskelet Disord*. 2018;19(1):1–9.
9. Nolan BM, Ankersen E, Wiater MJ. Reverse total shoulder arthroplasty improves function in cuff tear arthropathy. *Clin Orthop Relat Res*. 2011; 469(9):2476–82.
10. (OHSKO) OH and SC of O. Musculoskeletal Disorder (MSD) Prevention Guideline for Ontario. 2007;.
11. Cho CH, Song KS, Koo TW. Clinical outcomes and complications during the learning curve for reverse total shoulder arthroplasty: An analysis of the first 40 cases. *CiOs Clin Orthop Surg*. 2017;9(2):213–7.
12. Wilcox RB, Arslanian LE, Millett PJ. Rehabilitation following total shoulder arthroplasty. *J Orthop Sports Phys Ther*. 2005;35(12):821–36.
13. Roy J-S, MacDermid J, Woodhouse LJ. Measuring shoulder function : a systematic review of four questionnaires. *Arthritis Rheum*. 2009;61(5):623–32.
14. Richards RR, An KN, Bigliani LU, Friedman RJ, Gartsman GM, Cristina AG, et al. A standardized method for the assessment of shoulder function. *J Shoulder Elb Surg*. 1994;3(6):347–52.
15. Dawson J, Fitzpatrick R, Carr A. Questionnaire on the perceptions of patients about shoulder surgery. *J Bone Joint Surg (Br)*. 2018;78-B(4):593–600.
16. Leggin BG, Michener LA, Shaffer MA, Breneman SK, Iannotti JP, Williams GR. The Penn shoulder score: reliability and validity. *J Orthop Sports Phys Ther* [Internet]. 2006;36(3):138–51 Available from: <http://www.jospt.org/doi/abs/10.2519/jospt.2006.36.3.138%5Cnpapers3://publication/doi/10.2519/jospt.2006.36.3.138>.
17. Gazielly DF, Scarlat MM, Verborgt O. Long-term survival of the glenoid components in total shoulder replacement for arthritis. *Int Orthop*. 2014; 39(2):285–9.
18. Roe Y, Soberg HL, Bautz-holter E, Ostensjo S. A systematic review of measures of shoulder pain and functioning using the International classification of functioning , disability and health (ICF). *BMC Musculoskelet Disord*. 2013;14(1):1.
19. De Kleijn-De Vrankrijker MW. The international classification of impairments, disabilities, and handicaps (ICIDH): perspectives and developments (part i). *Disabil Rehabil*. 1995;17(3–4):109–11.
20. Karimi M, Brazier J. Health, health-related quality of life, and quality of life: what is the difference? *Pharmacoeconomics*. 2016;34(7):645–9.
21. WHO. Towards a Common Language for Functioning , Disability and Health ICF. WHO. 2002;1149:1–22 Available from: <http://www.who.int/classifications/icf/training/icfbeginnersguide.pdf>.
22. Ueda S, Okawa Y. The subjective dimension of functioning and disability: what is it and what is it for? *Disabil Rehabil*. 2003;25(11–12):596–601.
23. Fayed N, Kraus O, Elizabeth DEC, Peter K, Ankita R, Bostan C, et al. Generic patient-reported outcomes in child health research : a review of conceptual content using World Health Organization definitions. *Dev Med Child Neurol*. 2012;54:1085–95.
24. Bruesch A, Reynolds C, Hailey E, Martin J, Treadway L. Introduction,administration,scoring and generic version of the assessment. 2011;(December). Available from: [https://wcu.blackboard.com/bbcswebdav/institution/eBriefcase/sample\\_portfolio/RTHPortfolios/Lena\\_Treadway/content/e501alstonbrookmarketingproject.pdf%5Chttp, www.who.int/mental\\_health/media/en/76.pdf](https://wcu.blackboard.com/bbcswebdav/institution/eBriefcase/sample_portfolio/RTHPortfolios/Lena_Treadway/content/e501alstonbrookmarketingproject.pdf%5Chttp, www.who.int/mental_health/media/en/76.pdf).
25. Dreinhöfer K, Stucki G, Ewert T, Huber E, Ebenbichler G, Gutenbrunner C, et al. ICF Core sets for osteoarthritis. *J Rehabil Med Suppl*. 2004;44:75–80.
26. Cieza A, Brockow T, Ewert T, Amman E, Kollerits B, Chatterji S, et al. Linking health-status measurements to the international classification of functioning, Disability and Health. *J Rehabil Med*. 2002;34(5):205–10.
27. Hurst R. The international disability rights movement and the ICF. *Disabil Rehabil*. 2003;25(11–12):572–6.
28. Nordenfelt L. Action theory, disability and ICF. *Disabil Rehabil*. 2003;25(18): 1075–9.
29. Cieza A, Geyh S, Chatterji S, Kostanjsek N, Üstün B, Stucki G. ICF linking rules: An update based on lessons learned. *J Rehabil Med*. 2005;37(4):212–8.
30. Cieza A, Fayed N, Bickenbach J, Prodinger B. Refinements of the ICF linking rules to strengthen their potential for establishing comparability of health information. *Disabil Rehabil*. 2019;41(5):574–83.
31. Moher D, Liberati A, Tetzlaff J, Altman DG, Altman D, Antes G, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement (Chinese edition). *J Chinese Integr Med*. 2009;7(9):889–96.
32. MacDermid J. ICF Linkage Indicator Definitions. 2014;(2):1–2. Available from: [http://srs-mcmaster.ca/wp-content/uploads/2015/04/ICF-linkage-indicators\\_Final-to-Post.pdf](http://srs-mcmaster.ca/wp-content/uploads/2015/04/ICF-linkage-indicators_Final-to-Post.pdf).
33. Vincent JJ, Macdermid JC, King GJW, Grewal R. Linking of the patient rated elbow evaluation (PREE) and the American shoulder and elbow surgeons - elbow questionnaire (pASES-e) to the international classification of functioning disability and health (ICF) and hand core sets. *J Hand Ther*. 2015;28(1):61–8. <https://doi.org/10.1016/j.jht.2014.10.002>.
34. Alonazi WB, Thomas SA. Quality of Care and Quality of Life : Convergence or Divergence? 2014. p. 1–12.
35. Badcock LJ, Lewis M, Hay EM, McCarney R, Croft PR. Chronic shoulder pain in the community: a syndrome of disability or distress? *Ann Rheum Dis*. 2002;61(2):128–31.
36. Gilbert MK, Gerber C. Comparison of the subjective shoulder value and the constant score. *J Shoulder Elb Surg*. 2007;16(6):717–21.
37. Furtado, R., MacDermid, J.C., Bryant, D.M. et al. Interpretation and content validity of the items of the numeric rating version short-WORC to evaluate outcomes in management of rotator cuff pathology: a cognitive interview approach. *Health Qual Life Outcomes* 18, 88 (2020). <https://doi.org/10.1186/s12955-020-01339-7>.
38. Gobeze R, Denard PJ, Shishani Y, Romeo AA, Lederman E. Healing and functional outcome of a subscapularis peel repair with a stem-based repair after total shoulder arthroplasty. *J Shoulder Elb Surg*. 2017;26(9):1603–8. <https://doi.org/10.1016/j.jse.2017.02.013>.
39. Arumugam V, MacDermid JC, Grewal R. Content analysis of work limitation, Stanford Presenteeism, and work instability questionnaires using international classification of functioning, disability, and health and item perspective framework. *Rehabil Res Pract*. 2013;2013:1–11.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

