A systematic review and mapping exercise to assess the content validity of patient-reported outcome measures for adults having reconstructive surgery of the lower limb

From University of York, York, UK

Correspondence should be sent to J. Lister jennie.lister@york.ac.uk

Cite this article:
Bone Jt Open 2024;5(12):
1049–1066.

DOI: 10.1302/2633-1462.512.BJO-2024-0074.R1

J. Lister,[1] S. James,[1] H. K. Sharma,[2] C. Hewitt,[1] H. Fulbright,[3] H. Leggett,[1] C. McDaid[1]

[1]York Trials Unit (YTU), Department of Health Sciences, University of York, York, UK
[2]Department of Trauma & Orthopaedics, Hull Royal Infirmary, Hull, UK
[3]Centre for Reviews and Dissemination (CRD), University of York, York, UK

Aims

Lower limb reconstruction (LLR) has a profound impact on patients, affecting multiple areas of their lives. Many patient-reported outcome measures (PROMs) are employed to assess these impacts; however, there are concerns that they do not adequately capture all outcomes important to patients, and may lack content validity in this context. This review explored whether PROMs used with adults requiring, undergoing, or after undergoing LLR exhibited content validity and adequately captured outcomes considered relevant and important to patients.

Methods

A total of 37 PROMs were identified. Systematic searches were performed to retrieve content validity studies in the adult LLR population, and hand-searches used to find PROM development studies. Content validity assessments for each measure were performed following Consensus-based Standards for the selection of health measurement Instruments (COSMIN) guidelines. A mapping exercise compared all PROMs to a conceptual framework previously developed by the study team ('the PROLLIT framework') to explore whether each PROM covered important and relevant concepts.

Results

The systematic searches found 13 studies, while hand searches found 50 PROM development studies, and copies of all 37 measures. Although several studies discussed content validity, none were found which formally assessed this measurement property in the adult LLR population. Development of many PROMs was rated as inadequate, no PROM had sufficient content validity in the study population, and none covered all areas of the PROLLIT framework. The LIMB-Q was the most promising and comprehensive measure assessed, although further validation in a wider sample of LLR patients was recommended.

Conclusion

Current PROMs used in adults requiring, undergoing, or after undergoing LLR lack content validity and do not assess all important and relevant outcomes. There is an urgent need for improved outcome measurement in this population. This can be achieved through development of a new PROM, or through validation of existing measures in representative samples.

Take home message

- Outcome measurement for adult patients having lower limb reconstruction surgery is poor, and requires improvement.
- Good-quality patient-reported outcome measures are urgently needed to support clinicians in providing high-quality patient-centered care to this group.

Bone & Joint Open

Introduction

Patients undergoing lower limb reconstruction (LLR) can face a long and difficult treatment, recovery, and rehabilitation process. Adult patients report pain and loss of function, negative effects on work and employment, reduced ability to perform their usual hobbies and activities, psychological difficulties, and changes to their sense of self.[1-4] The impact of this process on patients can therefore be profound and felt across multiple and interacting areas of their lives.

Patient-reported outcome measures (PROMs) are a key component in assessing these kinds of impacts.[5] Many are currently used for adults requiring, undergoing, or after undergoing LLR, relating to areas such as limb function, pain, and health-related quality of life (HRQoL).[6-9] There are concerns, however, that these PROMs are not adequately or effectively capturing the full range of experiences important to this group; the lack of patient involvement in their development is a key issue as this is recognized as essential to the effective measurement of patient experience.[10]

A PROM which is not developed or validated in the target patient population, and/or which does not assess the full range of experiences important to that group, is likely to lack content validity. This describes the ability of a PROM to adequately reflect the construct it is aiming to measure. A PROM with good content validity is comprehensive, relevant, and comprehensible in the target population and context. Content validity is dependent on the population in which a measure is applied, and a PROM which exhibits good content validity in one setting may be inadequate in another. All other measurement properties can be affected by a lack of content validity, meaning it is widely considered to be the most important property of an effective PROM.[11] A PROM that is missing concepts important to a patient population (i.e. is not comprehensive), for example, risks having reduced validity and responsiveness when applied in that context. When a measure has validity, we can make statements and predictions about patients based on their scores.[12] For example, we may be able to predict differences in recovery or support requirements between patient groups and make plans for care based on that information. A lack of validity could therefore lead to inaccurate predictions and statements, which may result in poor quality care and potential wasting or misdirection of resources. For this reason, validity is also important when using outcome measures in research, for example to compare and/or assess the impact of interventions. In these cases, a lack of validity could lead to inaccurate comparisons between interventions, and incorrect conclusions regarding their effectiveness.

The *P*atient-*R*eported *O*utcome *M*easure for *L*ower *LI*mb recons*T*ruction (PROLLIT) study was designed in response to concerns about outcome measurement in the adult LLR population.[13] During Phase 1 of the study, a conceptual framework (Figure 1)[2] was developed through qualitative evidence synthesis,[1] and interviews with patients and orthopaedic healthcare professionals (HCPs)[2] to establish which outcomes were important to this group.

Study aims and objectives

This review formed Phase 2 of the PROLLIT study. We aimed to explore whether current PROMs used with the adult LLR population exhibit content validity and adequately capture patient experience. These aims were achieved in three stages: 1) a list of key PROMs was created based on previous peer-reviewed research in this area, and discussion with an expert panel of surgeons;[6,14,15] 2) the content validity of these PROMs was formally assessed, following the Consensus-based Standards for the selection of health measurement Instruments (COSMIN) guidelines,[16] which included carrying out systematic searches to retrieve existing content validity studies for each PROM; and 3) a conceptual mapping exercise was performed to explore whether each measure covered the concepts outlined in the PROLLIT framework as important to patients.

Findings were brought together to answer two key questions: 1) do key PROMs currently in use with adult LLR patients assess the outcomes identified as being important and relevant in Phase 1 of the PROLLIT study?; and 2) do these PROMs exhibit content validity in this patient population?

Methods

Selection of PROMs

In the first stage of the study, we collated information from three recent systematic reviews (2019 to 2023) which had explored outcome measurement in lower limb trauma and reconstruction to generate a list of PROMs currently in use for LLR patients.[6,14,15] The list was checked by the PROLLIT expert advisory panel of surgeons (n = 5) with expertise in orthopaedics and limb reconstruction, to ensure no key measures were missing. The resulting list included 37 key PROMs known to be currently in use for adults requiring, undergoing, or after undergoing LLR.

Descriptive data including PROM name, language, scope, target population, intended context of use, associated concepts or domains, number and types of items, recall period, scoring information, and time to complete was extracted for each PROM using a template created for the process. Table I shows key descriptive information for the included measures.

COSMIN assessment of content validity

In the second stage of the study, the 37 PROMs were assessed for content validity. This process was carried out following the COSMIN guidance for the assessment of the measurement properties of PROMs.[16] As per the guidance, systematic searches were first performed to identify any existing studies of content validity for the included PROMs. These were supplemented by hand searches to identify studies describing PROM development, and to collate copies of the measures.

Systematic search procedure

The systematic review was registered on PROSPERO on the 13 October 2023 (ref: CRD42023469835).

Eligibility criteria

The inclusion criteria was adult patients (aged 16+ years) requiring, undergoing, or after undergoing any/all types of reconstructive surgery for a lower limb condition. There was no limits on timescale following injury or condition onset. The exclusion criteria was patients aged under 16 years; patients requiring, undergoing, or after undergoing amputation of the lower limb; and patients requiring, undergoing, or after undergoing arthroplasty/joint replacement.

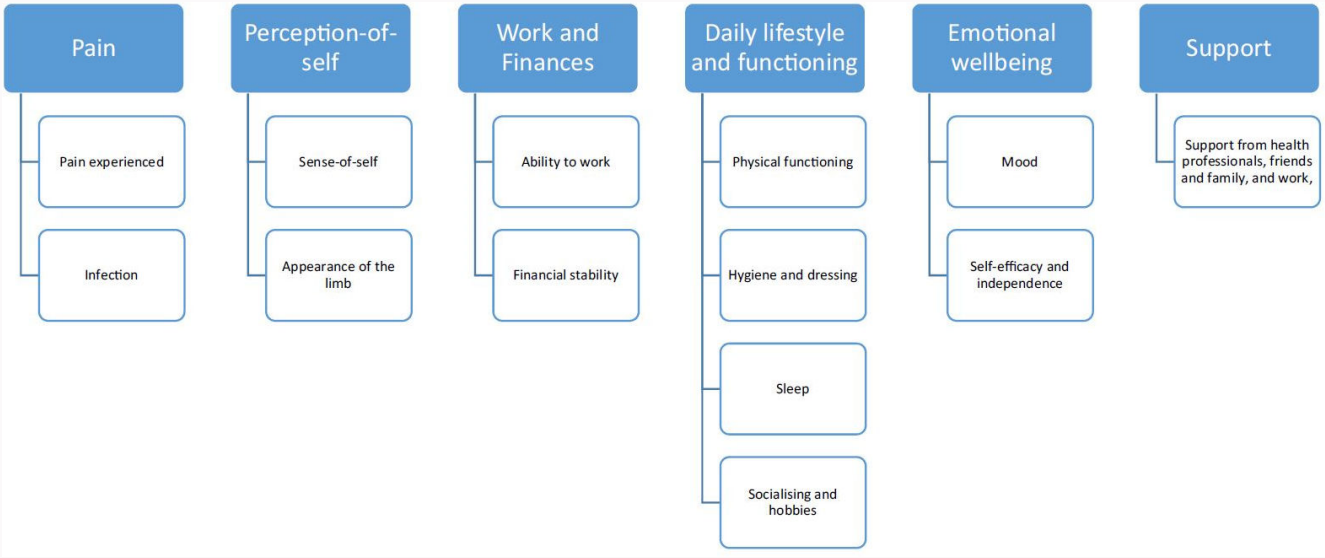

Fig. 1
The *P*atient-*R*eported *O*utcome *M*easure for *L*ower *LI*mb recons*T*ruction conceptual framework outlining what is important to adult lower limb reconstruction patients (reproduced from Leggett et al[2] under licence (CC BY 4.0)).

Outcome of interest

The outcome of interest was any of the 37 pre-identified PROMs, as outlined in Table I.

Types of studies

Studies were included if they were original, peer-reviewed studies where the content validity of a selected PROM has been assessed in the specified population (including professionals such as clinicians or researchers), or original, peer-reviewed studies detailing the development of each selected PROM.

Literature (including grey literature) detailing the content of each PROM instrument (e.g. user manuals, copies of PROMs) was also included.

Information sources and search strategy

An information specialist (HF) designed a preliminary search for MEDLINE, with input from the review team. The search strategy was designed to systematically identify all relevant studies on the measurement of the selected PROMs of interest in patients with lower limb conditions. The population terms on lower limb conditions were adapted from an earlier review conducted by the PROLLIT team.[1] Terwee et al's[65] methodological PubMed search filter for "finding studies on measurement properties of measurement instruments" was used. There were no restrictions on date or language applied to the searches. The search strategy was translated for use for the other bibliographic databases using relevant subject headings (controlled vocabularies) and search syntax, appropriate to each resource. A document detailing all search strategies as run is provided in the Supplementary Material.

The following sources were searched between 26 and 29 September 2023: MEDLINE(R) ALL (via Ovid); Embase (via Ovid); PsycINFO (via Ovid); Cumulated Index in Nursing and Allied Health Literature (CINAHL) Complete (via EBSCO); and Cochrane Central Register of Controlled Trials (CENTRAL) (via Wiley).

In addition, information on studies in progress, unpublished research, or research reported in the grey literature were sought by searching ClinicalTrials.gov and the World Health Organization International Clinical Trials Registry Platform (ICTRP).

Hand searches

Hand searches were conducted using Google Scholar and MEDLINE(R) (via Ovid) to identify development studies for each of the PROMs included in the study, along with copies of the measures. The COSMIN database was also searched to identify any existing PROM development assessments.[66]

Reference lists

Relevant systematic reviews were flagged during the screening process and reference lists searched to identify additional studies missed during the original electronic searches.

Screening

Following a short pilot screening process, a two-stage screening was carried out; title and abstract, followed by full-text screening. This was performed by two researchers independently (JL, SJ), using Covidence systematic review software (Veritas Health Innovation, Australia).

Content validity assessment

The COSMIN guidance and data extraction tables were used to formally assess content validity for each of the PROMs, with quality of evidence assessed using a modified GRADE approach.[16] The construct of interest was the PROLLIT conceptual framework, target population was adult LLR patients and context of use was the assessment of outcomes (i.e. evaluative). This was carried out by two independent researchers (JL, SJ), with disagreements resolved through discussion.

Table I. Descriptive information for all included patient-reported outcome measures.

PROM name	Acronym	Scope	Number of items	Type(s) of measure	Recall period(s)	Total score range and interpretation
Lower limb-specific						
American Academy of Orthopaedic Surgeons lower limb core scale[17]	AAOS-LLCS	Lower limb	7	Scales	Past week	0 to 100 Poor to best possible outcome
Foot and ankle ability measure[18]	FAAM	Leg, ankle, and foot	23 Optional 8-item subscale	Scales	Past week	0 to 100% of total possible score Lowest to highest level of functioning
Foot and Ankle Disability Index[19]	FADI	Foot and ankle	26 Optional 8-item subscale	Scales	Past week	0 to 100% of total possible score Lowest to highest level of functioning
Knee injury and Osteoarthritis Outcome Score[20,21]	KOOS	Knee	42	Scales	Present Past week	0 to 100 Extreme to no knee problems
Knee Society scoring system[22]	KSS	Knee (TKA)	30	Scales	Present	0 to 100 Worst to best function
Knee Society scoring system (short form)[23]	KSS SF	Knee (TKA)	10	Scales	Present	0 to 100 Worst to best function
The limb deformity-Scoliosis Research Society score[24]	LD-SRS	Lower limb	Section 1: 20 Section 2: 10	Scales	Present Last 6 months	1 to 5 Worst to best outcome
Lower Extremity Functional Scale[25]	LEFS	Lower limbs	20	Scales	Present	0 to 100% of total possible score Lower = greater disability
Patient-reported outcome instrument for lower extremity trauma[26-29]	LIMB-Q	Lower limbs	16 sections min = 6 items max = 15 items	Scales	Present Past week At time of surgery When last working	0 to 100 (calculated per section) Higher = better outcomes
Lysholm knee scale/score[30]	LKS	Knee	8	Scales	Present	0 to 100 Worst to best functioning
Olerund-Molander Ankle score[31]	OMAS	Ankle	9	Multiple choice	Present	0 to 100 Totally impaired to completely unimpaired
Stanmore Limb Reconstruction Score (in development: preliminary measure)[32]	SLRS	Lower limb	37	Scales	Present Past week Past 4 weeks	Not currently available
Tegner Activity Score[30]	TAS	Knee	1	Scale	Present	0 to 10 Higher = higher activity level
Toronto Extremity Salvage Score - Lower Limb[33]	TESS	Lower limbs	30	Scales	Past week	0 to 100 Higher = less disability
Visual analogue scale (foot and ankle)[34]	VAS-FA	Foot and ankle	20	VAS	Present	0 to 100

(*Continued*)

(*Continued*)

PROM name	Acronym	Scope	Number of items	Type(s) of measure	Recall period(s)	Total score range and interpretation
						Worst to best outcome
Other parts of the body						
Disabilities of the Arm, Shoulder and Hand questionnaire[35]	DASH	Upper limbs	30 Optional 4-item modules (x2)	Scales	Past week	0 to 100 Least to most disability
Musculoskeletal Function Assessment[7,36]	MFA	Upper and lower limbs	Dysfunction: 100 Bother: 10	Yes/No Scales	This week	0 to 100 (per subscale) Higher = more dysfunction/bother
Short Musculoskeletal Function Assessment[37]	SMFA	Upper and lower limbs	Dysfunction: 34 Bother: 12	Scales	This week	0 to 100 (per subscale) Higher = more dysfunction/bother
Oxford Hip Score[38,39]	OHS	Hip	12	Scales	Past 4 weeks	0 to 48 Higher = better outcomes
General health and wellbeing measures						
Brief pain inventory[40]	BPI	Clinical pain	9	Mixed	Past 24 hours	Severity: 0 to 40 Interference: 0 to 70 Higher = more severity/interference
Disability Rating Index[41]	DRI	Physical function	12	VAS	Present	0 to 100 Higher = more disability
EuroQol five-dimension five-level questionnaire[42]	EQ-5D-5L	HRQoL	6	Scales VAS	Present	5-digit health state 11111 = best possible health VAS: 0 to 100 Worst to best possible health
Frequency Intensity Time index[43]	FIT	Physical activity	3 parts	Scales	Not stated	0 to 100 Higher = more physically active
Nottingham Health Profile[44]	NHP	Perceived health	Part one: 38 Part two: 7	Yes/No	Present	Part one: 0 to 100 Part two: 0 to 7 Higher = greater number/severity of health problems
Patient Health Questionnaire - 9 item[45]	PHQ-9	Depression	9	Scales	Past 2 weeks	0 to 27 Higher = more severe depression
Short-Form 36 Health Survey[46,47]	SF-36	Functional health and wellbeing	36	Scales Yes/No	Present Past 4 weeks Compared to a year ago	0 to 100 (per subscale) Higher = better health/functioning
Short-Form 12 Health Survey[48]	SF-12	Functional health and wellbeing	12	Scales Yes/No	Present Past 4 weeks	Scores compared to mean of 50 and SD of 10

(*Continued*)

(*Continued*)

PROM name	Acronym	Scope	Number of items	Type(s) of measure	Recall period(s)	Total score range and interpretation
						> 50 = above average health < 50 = below average health
Veterans RAND 12-item health survey[49]	VR-12	Functional health and wellbeing	14	Scales	Present Past 4 weeks Compared to year ago	As SF-12
Sickness Impact Profile[50-54]	SIP	Perceived health	136	Yes/No	Present	0 to 100 Higher = more dysfunction
Visual analogue scale (pain)	VAS pain	Pain	1	VAS	Present	0 to 100 No pain to worst pain imaginable
Patient-Reported Outcomes Measurement Information System item banks[53,54]						
Ability to participate social roles/ activities[55]	PROMIS-APS	Participation in social roles	35	Scales	Present	Scoring tables used convert raw scores to t-scores Higher t-score= more of the outcome
Emotional distress (depression)[56-58]	PROMIS-D	Depression	28	Scales	Past 7 days	As above
Emotional distress (anxiety)[56]	PROMIS-A	Anxiety	29	Scales	Past 7 days	As above
Pain behaviour (v 2.0)[59]	PROMIS-PB	Pain behaviour	20	Scales	Past 7 days	As above
Pain interference (v 1.1)[60]	PROMIS-PI	Pain interference	40	Scales	Past 7 days	As above
Fatigue[61]	PROMIS-FIB	Fatigue	95	Scales	Past 7 days	As above
Physical functioning (Mobility v 2.1)[62-64]	PROMIS-PF mobility	Mobility	44	Scales	Present	As above

HRQoL, health-related quality of life; PROM, patient-reported outcome measure; TKA, total knee arthroplasty; VAS, visual analogue scale.

The ten COSMIN criteria for good content validity are shown in Figure 2.[16] A COSMIN content validity assessment answers these questions by combining information from the PROM development studies, content validity studies performed in the target population, and reviewer judgement. This evaluation was carried out for each PROM individually, in three stages:[16] 1) the quality of PROM development was assessed (where a rating for a PROM has been previously published, this was used); 2) where a content validity study was found in the systematic searches, its quality was assessed; and 3) overall content validity of each PROM was judged by combining information from stages 1 and 2 with researcher judgement of the comprehensiveness, comprehensibility, and relevance of the measure. For PROMs rated inadequate at the development stage or developed in a very different population, researcher judgement was prioritized.

Conceptual mapping

In the final stage of the study, a conceptual mapping exercise was carried out to compare each PROM to the PROLLIT framework. A table was created detailing each of the PROLLIT framework domains and sub-domains. For each PROM, the development study/studies and a copy of the measure were examined. Where a concept in the PROLLIT framework was covered by the PROM, the related item number(s) were recorded alongside a brief description. This procedure was carried out by two independent researchers (JL, SJ), with disagreements resolved through discussion.

Results

Literature search results

Results from the systematic and hand searches can be seen in Figure 3. Hand searches resulted in the collation of copies of all 37 PROMs along with 50 papers relating to their development, all of which were used in the PROM assessments. A preprint of a recent validation study for one of the PROMs was also provided by the study authors, resulting in 51 PROM development papers in total. The systematic searches identified 7,453 records, 7,206 of which were excluded at abstract stage, leaving 220 for full-text review. Of these, 13 were retained.

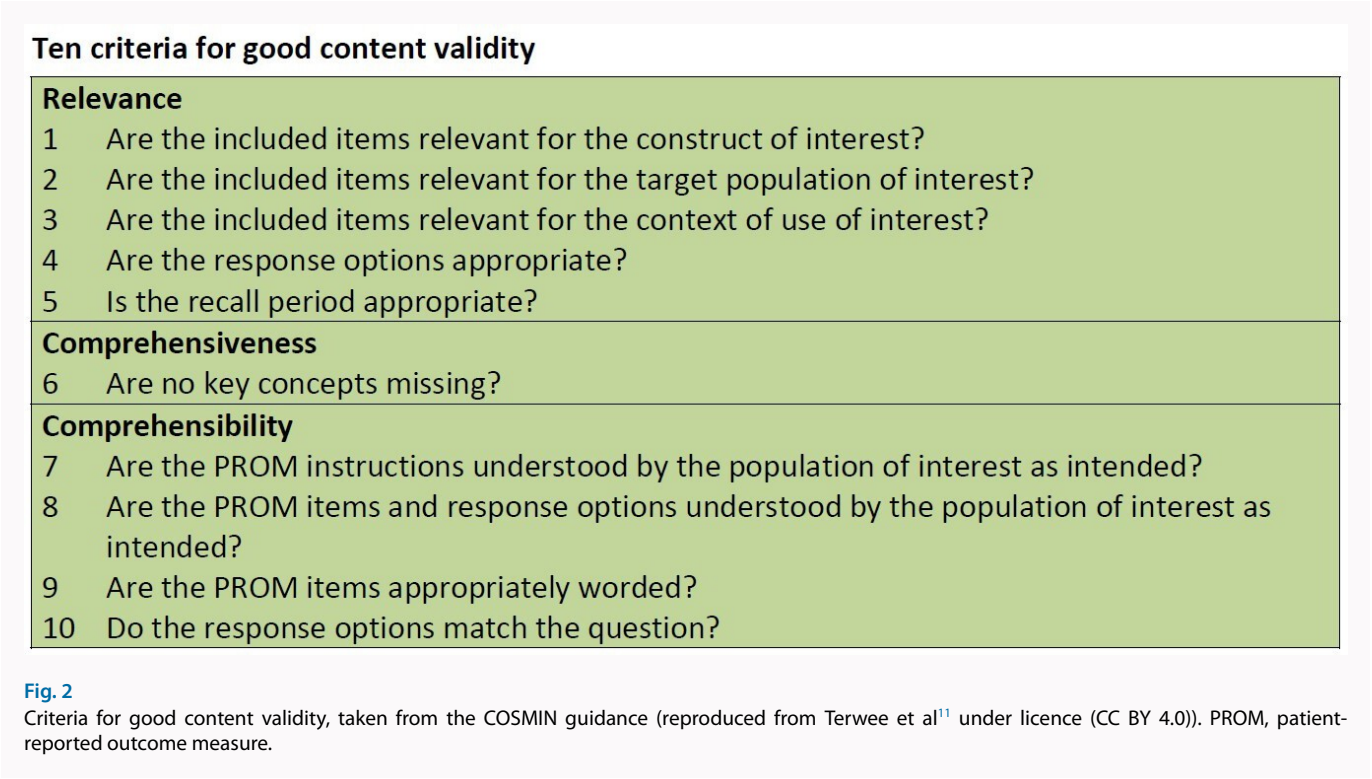

Fig. 2
Criteria for good content validity, taken from the COSMIN guidance (reproduced from Terwee et al[11] under licence (CC BY 4.0)). PROM, patient-reported outcome measure.

In total, therefore, 64 peer-reviewed studies and 37 copies of PROMs were included in the review. Further details of the studies can be seen in Table II.

Several PROM translation/cross-cultural adaption studies passed full-text review, as the authors mentioned assessment of content validity;[16] however, as shown in Table II, these did not include an assessment that could be used for COSMIN evaluation. Several content validity studies were also included which had been carried out prior to finalization of the associated PROM; these were considered a stage of PROM development and not as separate studies. Overall, therefore, no studies were found which formally assessed the content validity of a finalized PROM in the target population.

Content validity assessment results

Results from the COSMIN assessment can be seen in Table III. Where PROM subscales generated separate scores, they were rated and reported separately. As described, COSMIN content validity assessment is based on the PROM development, available content validity studies, and researcher judgement.[16] With the exception of the LIMB-Q, however, the development of all PROMs was inadequate. Where PROM development is found to be inadequate, the PROM development paper is then not considered in the final assessment of content validity of the measure. Given this and the lack of content validity studies available, assessments for all measures but the LIMB-Q were based on researcher judgement only and evidence quality was very low. Relevance and comprehensibility of PROMs was mixed, and none were considered comprehensive (i.e. none assessed all areas of the PROLLIT framework). Overall, most measures had indeterminate content validity ratings.

Conceptual mapping results

Table IV shows a summary of the conceptual mapping results. All PROMs covered at least one area of the PROLLIT framework, with physical function being the most commonly assessed (n = 31). The specificity of conceptual matches differed between PROMs. For example, general health and wellbeing measures referred more broadly to the effects of "health" (SF-36) or "present state of health" (NHP) on areas such as social life, ability to walk around, or ability to wash and dress. Lower-extremity measures, however, assessed concepts more specifically, for example difficulties walking related to problems with the leg. These measures were arguably of more relevance to the LLR population. No PROM assessed all areas of the PROLLIT framework, and the sub-domains of pain relating to infection, support from family and friends and support from work were not included in any. The LIMB-Q was the most comprehensive measure, while the LD-SRS, SLRS, MFA and NHP also covered a broad range of areas; these are discussed in more detail in the sections that follow.

Toronto Extremity Salvage Score – lower limb

The TESS had a rating of 'sufficient' for relevance and comprehensibility, and 'insufficient' for comprehensiveness, based on very low-quality evidence (reviewer judgement only). The measure was designed to evaluate physical function 'in the last week' following lower limb salvage due to sarcoma.[33] It was developed with upper- and lower-limb sarcoma patients, and therefore included a subset of the full LLR population. In the mapping exercise, the TESS covered the areas of pain experience, sense of self, ability to work, physical functioning, hygiene and dressing, and socializing and hobbies. Items referred to physical function only, for example relating to physical ability to participate in usual leisure activities. The TESS may hold promise as measure of

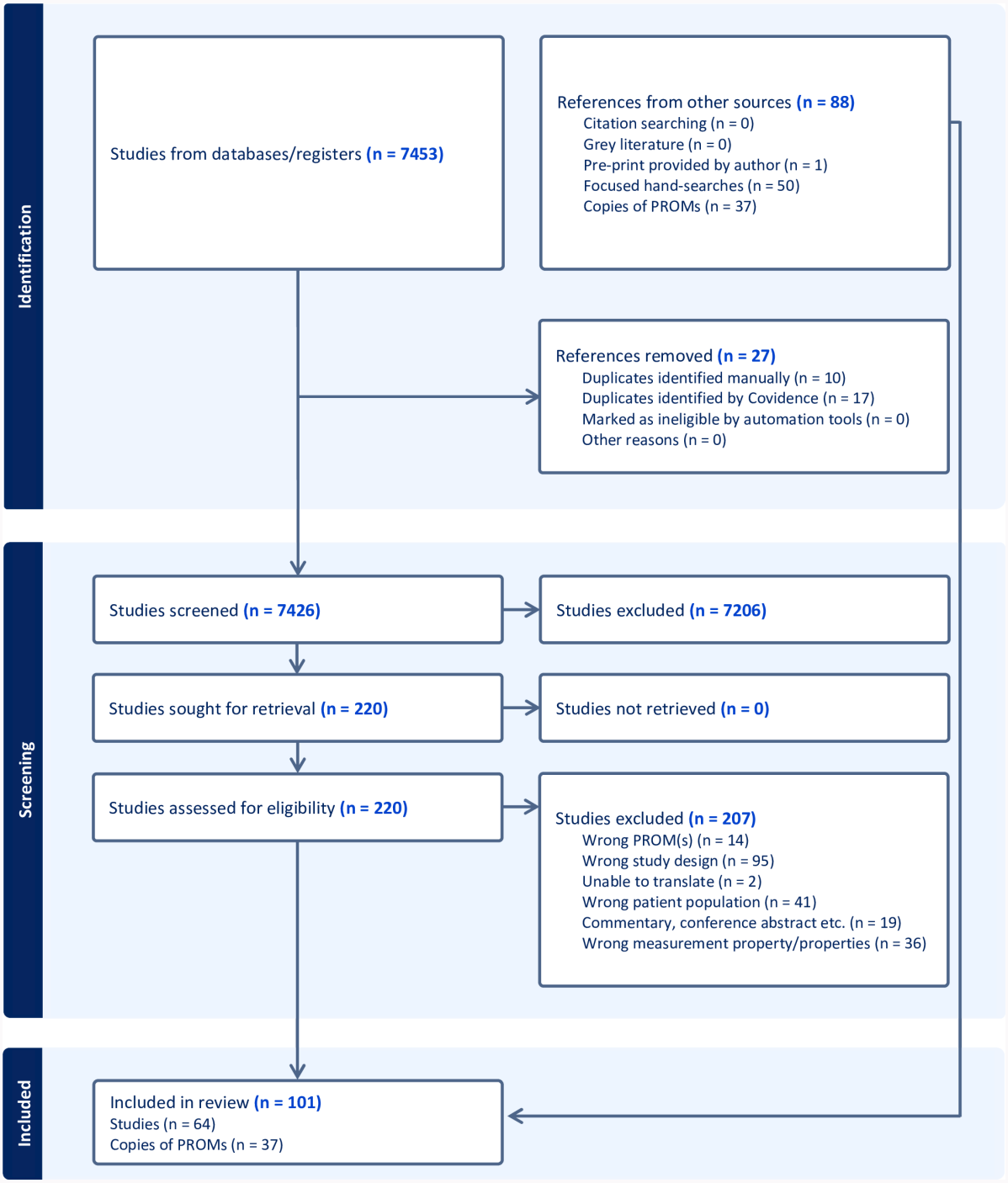

Fig. 3
A PRISMA-style flow diagram for the study searches. PROM, patient-reported outcome measure.

physical function following LLR but further validation in a wider and more representative sample is required.

The Limb Deformity-Scoliosis Research Society Score

The LD-SRS had ratings of 'indeterminate' for relevance and comprehensibility, and 'insufficient' for comprehensiveness, based on very low-quality evidence (reviewer judgement only). The measure was developed to assess the effects of limb deformity on HRQoL[24] for all relevant patients, and for those who have had treatment. In the mapping exercise, the LD-SRS assessed pain experience, sense of self, appearance of limb, ability to work, financial stability, physical functioning, socializing and hobbies, and mood. Content validity assessments were mixed, as the LD-SRS was designed for patients with deformity of any limb(s) and therefore a subset of the LLR population. Several items in particular were not considered relevant to patients having LLR due to injury to a previously 'normal' limb (e.g. "Compared with before treatment, how do you feel you now look?"). This measure may be valuable in assessing outcomes in some patients but was not comprehensive in relation to the PROLLIT framework.

Table II. Characteristics of the included studies.

Reference	PROM	Study type	Source/search method	Additional comments/study features
Johanson et al[17] (2004)	AAOS LLCS	Development	Systematic	
Kitaoka et al[67] (1994)	AOFAS	Development	Hand	
Charles and Cleeland[40] (2009)	BPI	Development	Hand	Development detailed in PROM user guide
Hudak et al[35] (1996)	DASH	Development	Hand	
Salén et al[41] (1994)	DRI	Development	Hand	
The EuroQol Group (1990)[42]	EQ-5D-5L	Development	Hand	
Martin et al[18] (2005)	FAAM	Development	Hand	
Martin et al[19] (1999)	FADI	Development	Hand	
Kasari[43] (1976)	FIT	Development	Hand	
Roos and Lohmander[20] (2003)	KOOS	Development	Hand	
Roos et al[21] (1998)	KOOS	Development	Hand	
Noble et al[22] (2012)	KSS	Development	Hand	
Scuderi et al[23] (2016)	KSS-SF	Development	Hand	
Fabricant et al[24] (2016)	LD-SRS	Development	Hand	
Binkley et al[25] (1999)	LEFS	Development	Hand	
Mundy et al[26] (2020)	LIMB-Q	Development	Systematic	
Mundy et al[28] (2020)	LIMB-Q	Development	Systematic	Content validity study as part of PROM development
Mundy et al[27] (2019)	LIMB-Q	Development	Systematic	
Mundy et al[29] (2024)	LIMB-Q	Development	LIMB-Q team	
Simonsen et al[68] (2023)	LIMB-Q	Cross-cultural adaptation/validation	Systematic	Adaptation and content validity which informed development of LIMB-Q
Tegner and Lysholm[30] (1985)	LKS/TAS	Development	Hand	
Engelberg et al[36] (1996)	MFA	Development	Hand	
Martin et al[7] (1996)	MFA	Development	Systematic	
Hunt et al[44] (1985)	NHP	Development	Hand	
Dawson et al[38] (1996)	OHS	Development	Hand	
Murray et al[39] (2007)	OHS/OKS	Development	Hand	
Olerud and Molander[31] (1984)	OMAS	Development	Hand	
Kroenke et al[45] (2001)	PHQ-9	Development	Hand	
Lai et al[61] (2011)	PROMIS fatigue	Development	Hand	
Ader[69] (2007)	PROMIS item banks	Development	Hand	
DeWalt et al[70] (2007)	PROMIS item banks	Development	Hand	
Castel et al[71] (2008)	PROMIS social health item banks	Development	Hand	
Hahn et al[55] (2010)	PROMIS-APS	Development	Hand	
Kelly et al[58] (2011)	PROMIS-D	Development	Hand	
Pilkonis et al[57] (2014)	PROMIS-D	Development	Hand	
Pilkonis et al[56] (2011)	PROMIS-D, PROMIS-A	Development	Hand	
Cook et al[72] (2013)	PROMIS-PB	Development	Hand	
Revicki et al[59] (2009)	PROMIS-PB	Development	Hand	
Hays et al[64] (2013)	PROMIS-PF	Development	Hand	

(*Continued*)

(*Continued*)

Reference	PROM	Study type	Source/search method	Additional comments/study features
Rose et al[62] (2008)	PROMIS-PF	Development	Hand	
Rose et al[63] (2014)	PROMIS-PF	Development	Hand	
Amtmann et al[60] (2010)	PROMIS-PI	Development	Hand	
Marshall and Hays[73] (1994)	PSQ-9	Development	Hand	
Ware et al[48] (1996)	SF-12	Development	Hand	
Ware et al[47] (1993)	SF-36	Development (user guide)	Hand	
Ware and Sherbourne[46] (1992)	SF-36	Development	Hand	
Bergner et al[50] (1981)	SIP	Development	Hand	
Bergner et al[51] (1976)	SIP	Development	Hand	
Bergner et al[52] (1976)	SIP	Development	Hand	
Gilson et al[53] (1975)	SIP	Development	Hand	
Pollard et al[54] (1976)	SIP	Development	Hand	
Wright et al[32] (2021)	SLRS	Development	Hand	
Scott et al[74] (2014)	SMFA	Validation in study population	Systematic	Term 'content validity' used, no discussion with patients or professionals
Swiontkowski et al[37] (1999)	SMFA	Development	Hand	
Kask et al[75] (2021)	TESS	Cross-cultural adaptation/validation	Systematic	Term 'content validity' used, mention of patient interviews, no reporting of discussion with patients or professionals
Ocaktan et al[76] (2021)	TESS	Cross-cultural adaptation/validation	Systematic	Participants asked about comprehensibility, no results reported, no formal content validity assessment
Srisawat et al (2018)[77]	TESS	Cross-cultural adaptation/validation	Systematic	Term 'content validity' used, no reporting of discussion with patients or professionals
Rossi et al[78] (2020)	TESS	Cross-cultural adaptation/validation	Systematic	Discussion with patients and cognitive debriefing, no testing of final measure and very limited reporting
Saebye et al[79] (2014)	TESS	Cross-cultural adaptation/validation	Systematic	Mention of comprehensibility checks with patients and professionals, no formal content validity assessment
Davis et al[33] (1996)	TESS lower limb	Development	Systematic	
Brokelman et al[80] (2012)	VAS	Development	Hand	
Richter et al[34] (2006)	VAS-FA	Development	Hand	
Iqbal et al[49] (2007)	VR-12	Development	Hand	

A, emotional distress (anxiety); AAOS LLCS, American Academy of Orthopaedic Surgeons lower limb core scale; AOFAS, American Orthopaedic Foot & Ankle Society; APS, ability to participate in social roles/activities; BPI, Brief Pain Inventory; D, emotional distress (depression); DASH, Disabilities of the Arm, Shoulder and Hand Questionnaire; DRI, Disability Rating Index; EQ-5D-5L, EuroQol five-dimension five-level questionnaire; FAAM, Foot and Ankle Ability Measure; FADI, Foot and Ankle Disability Index; FIT, Frequency Intensity Time index; KOOS, Knee injury and Osteoarthritis Outcome Score; KSS, Knee Society Score; KSS-SF, Knee Society Score (short form); LD-SRS, Limb Deformity-Scoliosis Research Society score; LEFS, Lower Extremity Functional Score; LKS, Lysholm Knee Score; MFA, Musculoskeletal Function Assessment; NHP, Nottingham Health Profile; OHS, Oxford Hip Score; OKS, Oxford Knee Score; OMAS, Olerud-Molander Ankle Score; PB, pain behaviour; PF, physical functioning; PHQ-9, Patient Health Questionnaire - 9 items; PI, pain interference; PROM, patient-reported outcome measure; PROMIS, Patient-Reported Outcome Measure Information System; SF-12, Short Form 12-item health survey; SF-36, short form 36-item health survey; SIP, Sickness Impact Profile; SLRS, Stanmore Limb Reconstruction Score; SMFA, Short Musculoskeletal Function Assessment; TAS, Tagner Activity Score; TESS, Toronto Extremity Salvage Score; VAS, visual analogue scale; VAS-FA, visual analogue scale (foot and ankle); VR-12, Veterans RAND 12-item health survey.

Stanmore Limb Reconstruction Score

As development of the SLRS is ongoing, COSMIN content validity assessment was not carried out at this stage. However, the measure was included in this review as it is one of a small number designed for the adult LLR population.[32] In the mapping exercise, the preliminary SLRS covered pain experience, appearance of limb, ability to work, physical functioning, hygiene and dressing, sleep, socializing and

Table III. Results of the COSMIN content validity assessment.

PROM	Quality of PROM development	Relevance rating* (+ / - / ± / ?)	Comprehensiveness rating* (+ / - / ± / ?)	Comprehensibility rating* (+ / - / ± / ?)	Overall content validity rating* (+ / - / ±)	Quality of evidence
Lower limb-specific						
AAOS LLCS	I	+	-	-	±	Very low
FAAM – ADL	I[81]	+	-	+	±	Very low
FAAM – sports	I[81]	±	-	±	±	Very low
FADI – ADL	I[81]	+	-	+	±	Very low
FADI – sports	I[81]	±	-	±	±	Very low
KOOS	I[82,83]	±	-	±	±	Very low
KSS	I	±	-	+	±	Very low
KSS-SF – symptoms	I	+	-	±	±	Very low
KSS-SF – satisfaction	I	±	-	+	±	Very low
KSS-SF – activities	I	±	-	-	±	Very low
LD-SRS – all patients	I	+	-	+	±	Very low
LD-SRS – post-treatment	I	±	-	±	±	Very low
LEFS	I[81,83,84]	+	-	+	±	Very low
LKS	I[83]	-	-	+	±	Very low
OMAS	I[81]	+	-	±	±	Very low
TAS	I[83]	-	-	-	-	Very low
TESS	I	+	-	+	±	Very low
VAS-FA	I[81]	-	-	-	-	Very low
LIMB-Q scales						
Appearance: reconstruction	A	+	±	+	±	Moderate
Appearance: amputation*	A	-	-	+	±	Moderate
Physical function	A	+	±	+	±	Moderate
Symptoms	A	±	±	+	±	Moderate
Expectations	A	+	±	+	±	Moderate
Financial impact	A	+	±	+	±	Moderate
Life impact	A	±	±	+	±	Moderate
Psychological	A	+	±	+	±	Moderate
Sexual	A	±	±	+	±	Moderate
Work	A	+	±	+	±	Moderate
Information	A	+	±	+	±	Moderate
Healthcare professional	A	±	±	+	±	Moderate
Office staff	A	±	±	+	±	Moderate
Treatment decision	A	±	±	+	±	Moderate
Prosthesis: function*	A	-	-	+	±	Moderate
Prosthesis: satisfaction*	A	-	-	+	±	Moderate
Other parts of the body						
DASH	I	-	-	+	±	Very low
MFA – dysfunction	I[81]	±	-	-	±	Very low
MFA – bother	I[81]	±	-	+	±	Very low

(*Continued*)

(*Continued*)

PROM	Quality of PROM development	Relevance rating* (+ / - / ± / ?)	Comprehensiveness rating* (+ / - / ± / ?)	Comprehensibility rating* (+ / - / ± / ?)	Overall content validity rating* (+ / - / ±)	Quality of evidence
SMFA – dysfunction	I[81]	±	-	±	±	Very low
SMFA – bother	I[81]	±	-	+	±	Very low
OHS	I	±	-	+	±	Very low
General health and wellbeing measures						
BPI	I[85]	+	-	+	±	Very low
DRI	I[86]	+	-		±	Very low
EQ-5D-5L	I[83,85–88]	+	-	+	±	Very low
NHP – part 1	I[87]	+	-	+	±	Very low
NHP – part 2	I[87]	±	-	+	±	Very low
PHQ-9	I[88]	+	-	+	±	Very low
SF-36 – general health	I[83,85–88]	-	-	+	±	Very low
SF-36 – physical functioning	I[83,85–88]	+	-	+	±	Very low
SF-36 – role physical	I[83,85–88]	+	-	+	±	Very low
SF-36 – role emotional	I[83,85–88]	+	-	+	±	Very low
SF-36 – social functioning	I[83,85–88]	+	-	±	±	Very low
SF-36 – bodily pain	I[83,85–88]	+	-	±	±	Very low
SF-36 – mental health and vitality	I[83,85–88]	+	-	+	±	Very low
SF-12	I[83,85–88]	+	-	+	±	Very low
SIP*	I	±	-	±	±	Very low
VR-12	I	±	-	+	±	Very low
PROMIS item banks (PROMIS-)						
APS	I	+	-	+	±	Very low
D	I	+	-	+	±	Very low
A	I	±	-	+	±	Very low
PB	I	±	-	+	±	Very low
PI	I	+	-	+	±	Very low
FIB	I	-	-	+	±	Very low
PF mobility	I[81]	+	-	+	±	Very low

Excluded from this table: SLRS (in development), VAS pain, and FIT (no development study available). Where PROM development rating was based on previous research, a citation is provided.

*(+ / - / ± / ?) (sufficient/insufficient/inconsistent/indeterminate).

A, emotional distress (anxiety); AAOS LLCS, American Academy of Orthopaedic Surgeons lower limb core scale; ADL, activities of daily living; APS, ability to participate in social roles/activities; BPI, Brief Pain Inventory; D, emotional distress (depression); DASH, Disabilities of the Arm, Shoulder and Hand Questionnaire; DRI, Disability Rating Index; EQ-5D-5L, EuroQol five-dimension five-level questionnaire; FAAM, Foot and Ankle Ability Measure; FADI, Foot and Ankle Disability Index; FIB, fatigue; KOOS, Knee injury and Osteoarthritis Outcome Score; KSS, Knee Society Score; KSS-SF, Knee Society Score (short form); LD-SRS, Limb Deformity-Scoliosis Research Society score; LEFS, Lower Extremity Functional Score; LKS, Lysholm Knee Score; MFA, Musculoskeletal Function Assessment; NHP, Nottingham Health Profile; OHS, Oxford Hip Score; OMAS, Olerud-Molander Ankle Score; PB, pain behaviour; PF, physical functioning; PHQ-9, Patient Health Questionnaire - 9 items; PI, pain interference; PROM, patient-reported outcome measure; PROMIS, Patient-Reported Outcome Measure Information System; SF-12, Short Form 12-item health survey; SF-36, short form 36-item health survey; SIP, Sickness Impact Profile; SMFA, Short Musculoskeletal Function Assessment; TAS, Tagner Activity Score; TESS, Toronto Extremity Salvage Score; VAS, visual analogue scale; VAS-FA, visual analogue scale (foot and ankle); VR-12, Veterans RAND 12-item health survey.

hobbies, and mood.[6] As this is a short measure, coverage was brief in some areas and the focus on a single surgical technique (external frame fixation) also meant that it was not relevant across the whole target population. Further assessment and validation of the SLRS is recommended once it is finalized.

Table IV. Overview of results from the conceptual framework mapping process.

PROM	Pain		Perception of self		Work and finances		Daily lifestyle and functioning				Emotional wellbeing		Support		
	Experienced	Infection	Sense of self	Appearance of limb	Ability to work	Financial stability	Physical functioning	Hygiene and dressing	Sleep	Socialising and hobbies	Mood	Self-efficacy and independence	HCP	Friends and family	Work
Lower limb-specific															
AAOS- LLCS	Y	-	-	-	-	-	Y	Y	-	-	-	-	-	-	-
FAAM	-	-	-	-	-	-	Y	Y	-	Y	-	-	-	-	-
FADI	Y	-	-	-	-	-	Y	-	Y	Y	-	-	-	-	-
KOOS	Y	-	-	-	-	-	Y	Y	-	Y	-	-	-	-	-
KSS	Y	-	-	-	-	-	Y	-	-	Y	-	-	-	-	-
KSS-SF	Y	-	-	-	-	-	Y	-	-	Y	-	-	-	-	-
LD-SRS	Y	-	Y	Y	Y	Y	Y	-	-	Y	Y	-	-	-	-
LEFS	-	-	-	-	-	-	Y	Y	-	Y	-	-	-	-	-
LIMB-Q	Y	-	-	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	-	-
LKS	Y	-	-	-	-	-	Y	-	-	-	-	-	-	-	-
OMAS	Y	-	-	-	Y	-	Y	-	-	-	-	-	-	-	-
SLRS	Y	-	-	Y	Y	-	Y	Y	Y	Y	Y	-	-	-	-
TAS	-	-	-	-	-	-	Y	-	-	-	-	-	-	-	-
TESS	Y	-	Y	-	Y	-	Y	Y	-	Y	-	-	-	-	-
VAS-FA	Y	-	-	-	Y	-	Y	-	-	-	-	-	-	-	-
Other parts of the body															
DASH	Y	-	-	-	Y	-	Y	Y	Y	Y	-	Y	-	-	-
MFA	-	-	Y	-	Y	-	Y	Y	Y	Y	Y	Y	-	-	-
SMFA	-	-	Y	-	Y	-	Y	Y	Y	Y	Y	Y	-	-	-
OHS	Y	-	-	-	Y	-	Y	Y	-	-	-	-	-	-	-
General health and wellbeing measures															
BPI	Y	-	-	-	Y	-	Y	-	Y	Y	Y	-	-	-	-

(*Continued*)

(*Continued*)

PROM	Pain		Perception of self		Work and finances		Daily lifestyle and functioning				Emotional wellbeing		Support		
	Experienced	Infection	Sense of self	Appearance of limb	Ability to work	Financial stability	Physical functioning	Hygiene and dressing	Sleep	Socialising and hobbies	Mood	Self-efficacy and independence	HCP	Friends and family	Work
DRI	-	-	-	-	-	-	Y	Y	-	Y	-	-	-	-	-
EQ-5D-5L	Y	-	-	-	Y	-	Y	Y	-	Y	Y	-	-	-	-
FIT	-	-	-	-	-	-	Y	-	-	-	-	-	-	-	-
NHP	Y	-	Y	-	Y	-	Y	Y	Y	Y	Y	Y	-	-	-
PHQ-9	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-
SF-36	Y	-	-	-	Y	-	Y	Y	-	Y	Y	-	-	-	-
SF-12	Y	-	-	-	Y	-	Y	-	-	Y	Y	-	-	-	-
SIP	-	-	Y	-	Y	-	Y	Y	-	Y	Y	-	-	-	-
VAS pain	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-
VR-12	Y	-	-	-	Y	-	Y	-	-	Y	Y	-	-	-	-
PROMIS item banks (PROMIS-)															
APS	-	-	-	-	Y	-	-	-	-	Y	-	-	-	-	-
D	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-
A	-	-	-	-	-	-	-	-	-	-	Y	-	-	-	-
PB	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-
PI	Y	-	-	-	Y	-	Y	-	-	Y	Y	-	-	-	-
FIB	-	-	-	-	Y	-	-	Y	-	Y	Y	-	-	-	-
PF mobility	-	-	-	-	-	-	Y	-	-	-	-	-	-	-	-

A, emotional distress (anxiety); AAOS-LLCS, American Academy of Orthopaedic Surgeons lower limb core scale; APS, ability to participate in social roles/activities; BPI, Brief Pain Inventory; D, emotional distress (depression); DASH, Disabilities of the Arm, Shoulder and Hand questionnaire; DRI, Disability Rating Index; EQ-5D-5L, Euroqol five-dimension five-level questionnaire; FAAM, Foot and Ankle Ability Measure; FADI, Foot and Ankle Disability Index; FIB, fatigue; FIT, Frequency Intensity Time index; HCP, healthcare professional; KOOS, Knee injury and Osteoarthritis Outcome Score; KSS (SF), Knee Society scoring system (short form); LD-SRS, Limb Deformity-Scoliosis Research Society score; LEFS, Lower Extremity Functional Scale; LKS, Lysholm Knee Scale/Score; NHP, Nottingham Health Profile; OHS, Oxford Hip Score; OMAS, Olerund-Molander Ankle Score; PB, Pain behaviour (v2.0); PF, physical functioning; PHQ-9, Patient Health Questionnaire - 9 item; PI, Pain interference (v1.1); PROMIS, Patient-Reported Outcomes Measurement Information System; SF-12, Short Form 12 Health Survey; SF- 36, Short Form 36 Health Survey; SIP, Sickness Impact Profile; SLRS, Stanmore Limb Reconstruction Score; (S)MFA, (Short) Musculoskeletal Function Assessment; TAS, Tegner Activity Score; TESS, Toronto Extremity Salvage Score - Lower Limb; VAS, visual analogue scale; VAS-FA, visual analogue scale (foot and ankle); VR-12, Veterans RAND 12 item health survey.

Musculoskeletal Function Assessment

The MFA has two subscales: the 'dysfunction' and 'bother' indices. The dysfunction index was rated 'indeterminate' for relevance and 'insufficient' for comprehensiveness and comprehensibility. The bother index was rated 'indeterminate' for relevance, 'insufficient' for comprehensiveness, and 'sufficient' for comprehensibility. Both subscales had 'indeterminate' content validity ratings overall, based on very low-quality evidence (reviewer judgement only). The MFA was designed to assess functioning 'this week' in patients with a variety of musculoskeletal disorders of the limbs.[7] In the mapping exercise, it covered sense of self, ability to work, physical functioning, hygiene and dressing, sleep, socializing and hobbies, mood, and self-efficacy and independence. However, as it was not designed for the LLR population many items related to 'injury or arthritis', and/or to 'hands and arms'. While elements of the MFA may be relevant to measuring outcomes in LLR patients, it would need further adaptation and validation to be recommended.

Nottingham Health Profile

The NHP has two subscales. Part 1 had 'sufficient' relevance and comprehensibility, and 'insufficient' comprehensiveness, while Part 2 had 'indeterminate' relevance, 'sufficient' comprehensibility, and 'insufficient' comprehensiveness. Overall ratings for both subscales were 'indeterminate', based on very low-quality evidence (reviewer judgement only). The NHP was designed to survey health problems in the general population and clinical settings and refers broadly to "present state of health".[44] In the mapping exercise, the NHP covered pain experience, sense of self, ability to work, physical functioning, hygiene and dressing, sleep, socializing and hobbies, mood, and self-efficacy and independence. Wording for these items was not directly relevant to LLR and the NHP would require further validation in this patient group before it could be recommended.

LIMB-Q

The LIMB-Q is a recently developed PROM for assessing outcomes in patients having reconstruction or amputation for limb-threatening lower limb trauma.[26-29] A review copy of the measure was provided upon request by the LIMB-Q team. The measure consists of 16 subscales with eight to 12 items, with each being scored separately (i.e. the LIMB-Q can be used in a modular fashion). Relevance, comprehensiveness, and comprehensibility ratings for each subscale can be seen in Table III. Overall content validity ratings were 'indeterminate', based on moderate-quality evidence (PROM development studies and reviewer judgement). However, the measure was well-developed according to COSMIN standards, being rated as 'adequate' due to the 'worst score counts' rule, but with the majority of areas rated 'very good'. As the LIMB-Q is newly developed, no content validity studies have yet been carried out using the finalized PROM, however a recent international study confirmed its validity and reliability in a group of lower limb trauma patients across the world.[29] In the mapping exercise, the LIMB-Q was the most comprehensive measure, assessing pain experience, appearance of limb, ability to work, financial stability, physical functioning, hygiene and dressing, sleep, socializing and hobbies, mood, self-efficacy and independence, and support from health professionals. Given these strengths, we believe it to be the most promising and relevant of the PROMs for patients having LLR. However, it was designed for outcome measurement following lower limb trauma, and the international nature of the development study meant only 9% of the validation sample were UK-based (n = 66). Additionally, while the measure covered many areas of the PROLLIT framework, it did not cover them all. Further validation of the LIMB-Q would therefore be needed in LLR patients with aetiologies other than traumatic injury, and with a larger sample of UK-based patients before it could be recommended for use in these contexts.

Discussion

We performed a content validity assessment and mapping exercise with 37 PROMs currently used in the adult LLR population to explore whether they adequately assess outcomes for these patients. Systematic searches found no content validity studies relating to finalized PROMs, most of which had not been developed in this population. Content validity is an essential feature of an effective PROM,[11] and our study therefore suggests that outcomes for LLR are not being adequately measured, supporting concerns previously highlighted by the PROLLIT team.[89] In the COSMIN assessment, no PROM was judged to have sufficient content validity across the target population. Nonetheless, the mapping exercise highlighted several measures that covered multiple areas of the PROLLIT framework, which were considered in more detail.

This was a comprehensive review and mapping exercise of a wide range of PROMs used in LLR. It is the first study to formally assess the content validity of PROMs used in this population using the COSMIN guidance. The mapping exercise was informed by high-quality research involving LLR patients,[1,2] which has resulted in a thorough evaluation of the relevance and comprehensiveness of the included PROMs, and provided a detailed picture of the current state of patient-reported outcome measurement in the adult LLR population.

The study had some limitations – a combination of previously published systematic reviews, along with consultation with expert HCPs, was used to identify PROMs for assessment, and it is possible that some relevant measures were missed from this study. Nonetheless, the 37 included PROMs allowed a comprehensive overview of outcome measurement for LLR. As no content validity studies were found, the COSMIN method was difficult to apply for measures developed in very different populations, and assessment for most PROMs relied on researcher judgement alone, meaning the quality of evidence was very low. This limitation, however, also speaks to the lack of tailored and relevant outcome measures available for these patients.

Recommendations for future research

Based on the findings of this study, three potential avenues for future research are proposed. Further exploratory work would be recommended to identify which of these is the most appropriate:

Validation of the LIMB-Q

The LIMB-Q is a well-designed measure, developed in a population which most closely matched our population of interest. As it is a new measure, no content validity studies have been carried out, however we believe it to be a

promising PROM for use in this population which is deserving of further consideration. The measure has some caveats that could be addressed in future research. The LIMB-Q was designed for a subset of the LLR population only (those with lower limb trauma) and included patients undergoing amputation. When mapped to the PROLLIT framework, the measure did not cover pain related to infection, sense of self, support from friends and family, or support from work, while hygiene and dressing were assessed only briefly. Finally, it was developed in the USA and validation was carried out with patients internationally, meaning only a small number of UK-based patients was included. For the LIMB-Q to be used effectively in assessment of outcomes across the LLR population, therefore, we would recommend further validation in samples of patients having LLR for reasons other than trauma. Further items or measures may also need to be included alongside the measure to assess PROLLIT framework concepts not already covered, and concepts specific to non-trauma-related conditions requiring LLR. Validation in a larger UK-based sample of LLR patients is also recommended.

Validation of existing measures

The lack of content validity studies for many PROMs used in the adult LLR population is concerning. It is likely that HCPs working in this area will continue to use many of these PROMs and therefore we would strongly recommend that validity and reliability studies are carried out in samples of adult LLR patients to ensure outcomes are being captured effectively.

Development of a new PROM

Development of a new PROM designed to capture outcomes for the full range of adult LLR patients in the UK may be appropriate. This would need to be developed with a diverse sample of patients having LLR for any/all reasons to ensure it was comprehensive, relevant, and comprehensible.

To conclude, current PROMs used in adults requiring, undergoing, or after undergoing LLR lack content validity and do not assess all important and relevant outcomes. Improved outcome measurement in this group is urgently required, which may be achieved through validation of existing measures in a representative patient sample, or through development of a new PROM.

Supplementary material

Detailed search strategies as run for the systematic review.

References

1. **Leggett H**, **Scantlebury A**, **Byrne A**, **et al**. Exploring what is important to patients with regards to quality of life after experiencing a lower limb reconstructive procedure: a qualitative evidence synthesis. *Health Qual Life Outcomes*. 2021;19(1):158.
2. **Leggett H**, **Scantlebury A**, **Hewitt C**, **Sharma H**, **McDaid C**, **PROLLIT Study Collaborators**. What is important to adults after lower limb reconstruction surgery: a conceptual framework. *Qual Life Res*. 2023;32(6):1671–1682.
3. **Nightingale J**, **Shu'an KL**, **Scammell BE**, **Leighton P**, **Ollivere BJ**. What is important to patients who are recovering from an open tibial fracture? A qualitative study. *Clin Orthop Relat Res*. 2022;480(2):263–272.
4. **Rees S**, **Tutton E**, **Achten J**, **Bruce J**, **Costa ML**. Patient experience of long-term recovery after open fracture of the lower limb: a qualitative study using interviews in a community setting. *BMJ Open*. 2019;9(10):e031261.
5. **Bull C**, **Teede H**, **Watson D**, **Callander EJ**. Selecting and implementing patient-reported outcome and experience measures to assess health system performance. *JAMA Health Forum*. 2022;3(4):e220326.
6. **Antonios T**, **Barker A**, **Ibrahim I**, **et al**. A systematic review of patient-reported outcome measures used in circular frame fixation. *Strat Trauma Limb Reconstr*. 2019;14(1):34–44.
7. **Martin DP**, **Engelberg R**, **Agel J**, **Snapp D**, **Swiontkowski MF**. Development of a musculoskeletal extremity health status instrument: the musculoskeletal function assessment instrument. *J Orthop Res*. 1996;14(2):173–181.
8. **Stanhope J**. Brief Pain Inventory review. *Occup Med (Lond)*. 2016;66(6):496–497.
9. **Stewart AL**, **Hays RD**, **Ware JE**. The MOS short-form general health survey. Reliability and validity in a patient population. *Med Care*. 1988;26(7):724–735.
10. **Wiering B**, **de Boer D**, **Delnoij D**. Patient involvement in the development of patient-reported outcome measures: a scoping review. *Health Expect*. 2017;20(1):11–23.
11. **Terwee CB**, **Prinsen CAC**, **Chiarotto A**, **et al**. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Qual Life Res*. 2018;27(5):1159–1170.
12. **Streiner DL**, **Norman GR**, **Cairney J**. *Health Measurement Scales: A Practical Guide to Their Development and Use*. 5th ed. Norman GR, Cairney J, NormanGR, eds. Oxford: Oxford University Press, 2015.
13. **McDaid C**, **Sharma H**, **Leggett H**, **Scantlebury A**, **Hewitt C**. Measuring quality of life in patients undergoing reconstructive procedures: an overview of the PROLLIT study. *Orthop Proc*. 2021;103-B(SUPP_6).
14. **Aquilina AL**, **Claireaux H**, **Aquilina CO**, **et al**. What outcomes have been reported on patients following open lower limb fracture, and how have they been measured? *Bone Joint Res*. 2023;12(2):138–146.
15. **Besmens IS**, **Zoller FE**, **Guidi M**, **Giovanoli P**, **Calcagni M**. How to measure success in lower extremity reconstruction, which outcome measurements do we use a systematic review and metanalysis. *J Plast Surg Hand Surg*. 2023;57(1–6):505–532.
16. **Terwee CB**, **Prinsen C**, **Chiarotto A**, **De Vet H**, **Bouter LM**, **Alonso J**. *COSMIN Methodology for Assessing the Content Validity of PROMs–User Manual*. Amsterdam: VU University Medical Center, 2018.
17. **Johanson NA**, **Liang MH**, **Daltroy L**, **Rudicel S**, **Richmond J**. American Academy of Orthopaedic Surgeons lower limb outcomes assessment instruments. Reliability, validity, and sensitivity to change. *J Bone Joint Surg Am*. 2004;86-A(5):902–909.
18. **Martin RL**, **Irrgang JJ**, **Burdett RG**, **Conti SF**, **Van Swearingen JM**. Evidence of validity for the Foot and Ankle Ability Measure (FAAM). *Foot Ankle Int*. 2005;26(11):968–983.
19. **Martin R**, **Burdett R**, **Irrgang J**. Development of the foot and ankle disability index (FADI). *J Orthop Sports Phys Ther*. 1999;29(1):A32–A33.
20. **Roos EM**, **Lohmander LS**. The Knee injury and Osteoarthritis Outcome Score (KOOS): from joint injury to osteoarthritis. *Health Qual Life Outcomes*. 2003;1(1):1–8.
21. **Roos EM**, **Roos HP**, **Lohmander LS**, **Ekdahl C**, **Beynnon BD**. Knee Injury and Osteoarthritis Outcome Score (KOOS)--development of a self-administered outcome measure. *J Orthop Sports Phys Ther*. 1998;28(2):88–96.
22. **Noble PC**, **Scuderi GR**, **Brekke AC**, **et al**. Development of a new Knee Society scoring system. *Clin Orthop Relat Res*. 2012;470(1):20–32.
23. **Scuderi GR**, **Sikorskii A**, **Bourne RB**, **Lonner JH**, **Benjamin JB**, **Noble PC**. The knee society short form reduces respondent burden in the assessment of patient-reported outcomes. *Clin Orthop Relat Res*. 2016;474(1):134–142.
24. **Fabricant PD**, **Borst EW**, **Green SA**, **Marx RG**, **Fragomen AT**, **Rozbruch SR**. Validation of a modified Scoliosis Research Society instrument for patients with limb deformity: the limb deformity-Scoliosis Research Society (LD-SRS) score. *J Limb Lengthen Reconstr*. 2016;2(2):86.
25. **Binkley JM**, **Stratford PW**, **Lott SA**, **Riddle DL**, **Network NAORR**. The Lower Extremity Functional Scale (LEFS): scale development, measurement properties, and clinical application. *Phys Ther*. 1999;79(4):371–383.
26. **Mundy LR**, **Klassen A**, **Grier AJ**, **et al**. Identifying factors most important to lower extremity trauma patients: key concepts from the development of a patient-reported outcome instrument for lower extremity trauma, the LIMB-Q. *Plast Reconstr Surg*. 2020;145(5):1292–1301.

27. **Mundy LR**, **Klassen A**, **Grier J**, **et al**. Development of a patient-reported outcome instrument for patients with severe lower extremity trauma (LIMB-Q): protocol for a multiphase mixed methods study. *JMIR Res Protoc*. 2019;8(10):e14397.
28. **Mundy LR**, **Klassen A**, **Sergesketter AR**, **et al**. Content validity of the LIMB-Q: a patient-reported outcome instrument for lower extremity trauma patients. *J Reconstr Microsurg*. 2020;36(9):625–633.
29. **Mundy LR**, **Klassen AF**, **Pusic AL**, **et al**. The LIMB-Q: reliability and validity of a novel patient-reported outcome measure for lower extremity trauma patients. *Plast Reconstr Surg*. 2024.
30. **Tegner Y**, **Lysholm J**. Rating systems in the evaluation of knee ligament injuries. *Clin Orthop Relat Res*. 1985;198:43–49.
31. **Olerud C**, **Molander H**. A scoring scale for symptom evaluation after ankle fracture. *Arch Orth Traum Surg*. 1984;103(3):190–194.
32. **Wright J**, **Timms A**, **Fugazzotto S**, **Goodier D**, **Calder P**. Development of a patient-reported outcome measure in limb reconstruction: a pilot study assessing face validity. *Bone Jt Open*. 2021;2(9):705–709.
33. **Davis AM**, **Wright JG**, **Williams JI**, **Bombardier C**, **Griffin A**, **Bell RS**. Development of a measure of physical function for patients with bone and soft tissue sarcoma. *Qual Life Res*. 1996;5(5):508–516.
34. **Richter M**, **Zech S**, **Geerling J**, **Frink M**, **Knobloch K**, **Krettek C**. A new foot and ankle outcome score: questionnaire based, subjective, visual-analogue-scale, validated and computerized. *Foot Ankle Surg*. 2006;12(4): 191–199.
35. **Hudak PL**, **Amadio PC**, **Bombardier C**, **et al**. Development of an upper extremity outcome measure: the DASH (disabilities of the arm, shoulder, and head). *Am J Ind Med*. 1996;29(6):602–608.
36. **Engelberg R**, **Martin DP**, **Agel J**, **Obremsky W**, **Coronado G**, **Swiontkowski MF**. Musculoskeletal Function Assessment instrument: criterion and construct validity. *J Orthop Res*. 1996;14(2):182–192.
37. **Swiontkowski MF**, **Engelberg R**, **Martin DP**, **Agel J**. Short musculoskeletal function assessment questionnaire: validity, reliability, and responsiveness. *J Bone Joint Surg Am*. 1999;81-A(9):1245–1260.
38. **Dawson J**, **Fitzpatrick R**, **Carr A**, **Murray D**. Questionnaire on the perceptions of patients about total hip replacement. *J Bone Joint Surg Br*. 1996;78-B(2):185–190.
39. **Murray DW**, **Fitzpatrick R**, **Rogers K**, **et al**. The use of the Oxford hip and knee scores. *J Bone Joint Surg Br*. 2007;89-B(8):1010–1014.
40. **Charles S**, **Cleeland P**. *The Brief Pain Inventory. User Guide*. Houston, Texas: The University of Texas M. D. Anderson Cancer Center, 2009.
41. **Salén BA**, **Spangfort EV**, **Nygren AL**, **Nordemar R**. The disability rating index: an instrument for the assessment of disability in clinical settings. *J Clin Epidemiol*. 1994;47(12):1423–1435.
42. EuroQol - a new facility for the measurement of health-related quality of life. *Health Policy*. 1990;16(3):199–208.
43. **Kasari D**. Effects of Exercise and Fitness on Serum Lipids in College Women, unpublished Master's Thesis University of Montana. 1976.
44. **Hunt SM**, **McEwen J**, **McKenna SP**. Measuring health status: a new tool for clinicians and epidemiologists. *J R Coll Gen Pract*. 1985;35(273):185–188.
45. **Kroenke K**, **Spitzer RL**, **Williams JB**. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med*. 2001;16(9):606–613.
46. **Ware JE**, **Sherbourne CD**. The MOS 36-Item Short-Form Health Survey (SF-36). *Med Care*. 1992;30(6):473–483.
47. **Ware JE**, **Snow KK**, **Kosinski M**, **Gandek B**. *SF-36 Health Survey. Manual and Interpretation Guide*. Boston: The Health Institute, New England Medical Center, 1993: 10–16.
48. **Ware JE**, **Kosinski M**, **Keller SD**. A 12-item short-form health survey: construction of scales and preliminary tests of reliability and validity. *Med Care*. 1996;34(3):220–233.
49. **Iqbal SU**, **Rogers W**, **Selim A**, **Qian S**, **Lee A**, **Ren X**, **et al**. *The Veterans RAND 12 Item Health Survey (VR-12): What It Is and How It Is Used*. CHQOERs VA Medical Center Bedford, MA: CAPP Boston University School of Public Health, 2007: 1–12.
50. **Bergner M**, **Bobbitt RA**, **Carter WB**, **Gilson BS**. The sickness impact profile: development and final revision of a health status measure. *Med Care*. 1981;19(8):787–805.
51. **Bergner M**, **Bobbitt RA**, **Kressel S**, **Pollard WE**, **Gilson BS**, **Morris JR**. The sickness impact profile: conceptual formulation and methodology for the development of a health status measure. *Int J Health Serv*. 1976; 6(3):393–415.
52. **Bergner M**, **Bobbitt RA**, **Pollard WE**, **Martin DP**, **Gilson BS**. The sickness impact profile: validation of a health status measure. *Med Care*. 1976;14(1):57–67.
53. **Gilson BS**, **Gilson JS**, **Bergner M**, **et al**. The sickness impact profile. Development of an outcome measure of health care. *Am J Public Health*. 1975;65(12):1304–1310.
54. **Pollard WE**, **Bobbitt RA**, **Bergner M**, **Martin DP**, **Gilson BS**. The sickness impact profile: reliability of a health status measure. *Med Care*. 1976;14(2):146–155.
55. **Hahn EA**, **Devellis RF**, **Bode RK**, **et al**. Measuring social health in the patient-reported outcomes measurement information system (PROMIS): item bank development and testing. *Qual Life Res*. 2010;19(7):1035–1044.
56. **Pilkonis PA**, **Choi SW**, **Reise SP**, **et al**. Item banks for measuring emotional distress from the Patient-Reported Outcomes Measurement Information System (PROMIS®): depression, anxiety, and anger. *Assessment*. 2011;18(3):263–283.
57. **Pilkonis PA**, **Yu L**, **Dodds NE**, **Johnston KL**, **Maihoefer CC**, **Lawrence SM**. Validation of the depression item bank from the Patient-Reported Outcomes Measurement Information System (PROMIS) in a three-month observational study. *J Psychiatr Res*. 2014;56:112–119.
58. **Kelly MAR**, **Morse JQ**, **Stover A**, **et al**. Describing depression: congruence between patient experiences and clinical assessments. *Br J Clin Psychol*. 2011;50(1):46–66.
59. **Revicki DA**, **Chen W-H**, **Harnam N**, **et al**. Development and psychometric analysis of the PROMIS pain behavior item bank. *Pain*. 2009; 146(1–2):158–169.
60. **Amtmann D**, **Cook KF**, **Jensen MP**, **et al**. Development of a PROMIS item bank to measure pain interference. *Pain*. 2010;150(1):173–182.
61. **Lai J-S**, **Cella D**, **Choi S**, **et al**. How item banks and their application can influence measurement practice in rehabilitation medicine: a PROMIS fatigue item bank example. *Arch Phys Med Rehabil*. 2011;92(10 Suppl): S20–7.
62. **Rose M**, **Bjorner JB**, **Becker J**, **Fries JF**, **Ware JE**. Evaluation of a preliminary physical function item bank supported the expected advantages of the Patient-Reported Outcomes Measurement Information System (PROMIS). *J Clin Epidemiol*. 2008;61(1):17–33.
63. **Rose M**, **Bjorner JB**, **Gandek B**, **Bruce B**, **Fries JF**, **Ware JE**. The PROMIS Physical Function item bank was calibrated to a standardized metric and shown to improve measurement efficiency. *J Clin Epidemiol*. 2014;67(5): 516–526.
64. **Hays RD**, **Spritzer KL**, **Amtmann D**, **et al**. Upper-extremity and mobility subdomains from the Patient-Reported Outcomes Measurement Information System (PROMIS) adult physical functioning item bank. *Arch Phys Med Rehabil*. 2013;94(11):2291–2296.
65. **Terwee CB**, **Jansma EP**, **Riphagen II**, **de Vet HCW**. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Qual Life Res*. 2009;18(8): 1115–1123.
66. **COSMIN**. COSMIN: Database of systematic reviews of outcome measurement instruments. University Library VU Amsterdam. 2024. https://database.cosmin.nl/ (date last accessed 7 November 2024).
67. **Kitaoka HB**, **Alexander IJ**, **Adelaar RS**, **Nunley JA**, **Myerson MS**, **Sanders M**. Clinical rating systems for the ankle-hindfoot, midfoot, hallux, and lesser toes. *Foot Ankle Int*. 1994;15(7):349–353.
68. **Simonsen NV**, **Rölfing JD**, **Mundy LR**, **et al**. Danish translation and linguistic validation of the LIMB-Q, a PROM for traumatic lower LIMB injuries and amputations. *Eur J Plast Surg*. 2023;46(6):1255–1264.
69. **Ader DN**. Developing the Patient-Reported Outcomes Measurement Information System (PROMIS). *Med Care*. 2007;45(5):S1–S2.
70. **DeWalt DA**, **Rothrock N**, **Yount S**, **Stone AA**, **Group PC**. Evaluation of item candidates: the PROMIS qualitative item review. *Med Care*. 2007; 45(5 Suppl 1):S12–21.
71. **Castel LD**, **Williams KA**, **Bosworth HB**, **et al**. Content validity in the PROMIS social-health domain: a qualitative analysis of focus-group data. *Qual Life Res*. 2008;17(5):737–749.
72. **Cook KF**, **Keefe F**, **Jensen MP**, **et al**. Development and validation of a new self-report measure of pain behaviors. *Pain*. 2013;154(12):2867–2876.
73. **Marshall GN**, **Hays RD**. *The Patient Satisfaction Questionnaire Short-Form (PSQ-18)*. CA: Rand Santa Monica, 1994.
74. **Scott DJ**, **Watson JDB**, **Heafner TA**, **Clemens MS**, **Propper BW**, **Arthurs ZM**. Validation of the Short Musculoskeletal Function

Assessment in patients with battlefield-related extremity vascular injuries. *J Vasc Surg*. 2014;60(6):1620–1626.
75. **Kask G**, **Uimonen MM**, **Barner-Rasmussen I**, **Tukiainen EJ**, **Blomqvist C**, **Repo JP**. Further validation of the Toronto extremity salvage score for lower extremity soft tissue sarcoma based on Finnish patients. *J Plast Reconstr Aesthet Surg*. 2021;74(1):71–78.
76. **Ocaktan B**, **Deveci MA**, **Yapar A**, **Tokgöz MA**, **Şimşek A**. Cross-cultural adaptation and validation of the Turkish version of the Toronto Extremity Salvage Score in patients with musculoskeletal tumors. *Acta Orthop Traumatol Turc*. 2021;55(5):428–434.
77. **Srisawat P**, **Pairojboriboon S**, **Veeraphun P**, **Punyaratabandhu T**, **Chareonvareekul S**, **Songpatanasilp T**. Development, validation, and reliability of functional outcome questionnaire on toronto extremity salvage score [TESS] to thai version for patients with bone and soft-tissue sarcoma. *J Med Assoc Thailand*. 2018;101(1):97–102.
78. **Rossi L**, **Boffano M**, **Comandone A**, **et al**. Validation process of Toronto Exremity Salvage Score in Italian: a quality of life measure for patients with extremity bone and soft tissue tumors. *J Surg Oncol*. 2020;121(4):630–637.
79. **Sæbye C**, **Safwat A**, **Kaa AK**, **Pedersen NA**, **Keller J**. Validation of a Danish version of the Toronto Extremity Salvage Score questionnaire for patients with sarcoma in the extremities. *Dan Med J*. 2014;61(1):A4734.
80. **Brokelman RBG**, **Haverkamp D**, **van Loon C**, **Hol A**, **van Kampen A**, **Veth R**. The validation of the visual analogue scale for patient satisfaction after total hip arthroplasty. *Eur Orthop Traumatol*. 2012;3(2):101–105.
81. **Nguyen MQ**, **Dalen I**, **Iversen MM**, **Harboe K**, **Paulsen A**. Ankle fractures: a systematic review of patient-reported outcome measures and their measurement properties. *Qual Life Res*. 2023;32(1):27–45.
82. **Hansen CF**, **Jensen J**, **Odgaard A**, **et al**. Four of five frequently used orthopedic PROMs possess inadequate content validity: a COSMIN evaluation of the mHHS, HAGOS, IKDC-SKF, KOOS and KNEES-ACL. *Knee Surg Sports Traumatol Arthrosc*. 2022;30(11):3602–3615.
83. **Sabah SA**, **Hedge EA**, **Abram SGF**, **Alvand A**, **Price AJ**, **Hopewell S**. Patient-reported outcome measures following revision knee replacement: a review of PROM instrument utilisation and measurement properties using the COSMIN checklist. *BMJ Open*. 2021;11(10):e046169.
84. **Ratter J**, **Pellekooren S**, **Wiertsema S**, **et al**. Content validity and measurement properties of the Lower Extremity Functional Scale in patients with fractures of the lower extremities: a systematic review. *J Patient Rep Outcomes*. 2022;6(1):11.
85. **Ghai V**, **Subramanian V**, **Jan H**, **Doumouchtsis SK**, **CHORUS: An International Collaboration Harmonising Outcomes, Research, and Standards in Urogynaecology and Womens Health**. A systematic review highlighting poor quality of evidence for content validity of quality of life instruments in female chronic pelvic pain. *J Clin Epidemiol*. 2022;149:1–11.
86. **Craxford S**, **Deacon C**, **Myint Y**, **Ollivere B**. Assessing outcome measures used after rib fracture: a COSMIN systematic review. *Injury*. 2019;50(11):1816–1825.
87. **Chiarotto A**, **Terwee CB**, **Kamper SJ**, **Boers M**, **Ostelo RW**. Evidence on the measurement properties of health-related quality of life instruments is largely missing in patients with low back pain: a systematic review. *J Clin Epidemiol*. 2018;102:23–37.
88. **Powell PA**, **Carlton J**, **Woods HB**, **Mazzone P**. Measuring quality of life in Duchenne muscular dystrophy: a systematic review of the content and structural validity of commonly used instruments. *Health Qual Life Outcomes*. 2020;18(1):263.
89. **Leggett H**, **Scantlebury A**, **Sharma H**, **et al**. Quality of life following a lower limb reconstructive procedure: a protocol for the development of a conceptual framework. *BMJ Open*. 2020;10(12):e040378.

Author information

J. Lister, BSc, MSc, PhD, Research Fellow
S. James, BSc, MSc, Research Fellow
C. Hewitt, BSc, MSc, PhD, Director, York Trials Unit
H. Leggett, BSc, MSc, PhD, Research Fellow
C. McDaid, BSc, MSc, PhD, Professor of Applied Health Research
York Trials Unit (YTU), Department of Health Sciences, University of York, York, UK.

H. K. Sharma, MB BS, MS (Orth), FRCS (Tr&Orth), FRCS (London), M Ch (Orth), FRCS (Glas), Orthopaedic Surgeon, Department of Trauma & Orthopaedics, Hull Royal Infirmary, Hull, UK.

H. Fulbright, MCLIP, BA Hons, PGDip LIS, MA, PhD, Research Fellow in Information Science, Information Specialist, Centre for Reviews and Dissemination (CRD), University of York, York, UK.

Author contributions

J. Lister: Conceptualization, Formal analysis, Investigation, Project administration, Writing – original draft, Writing – review & editing.
S. James: Formal analysis, Writing – review & editing.
H. K. Sharma: Conceptualization, Supervision, Writing – review & editing.
C. Hewitt: Conceptualization, Supervision, Writing – review & editing.
H. Fulbright: Methodology, Writing – review & editing, Investigation, Resources.
H. Leggett: Conceptualization, Writing – review & editing.
C. McDaid: Conceptualization, Supervision, Writing – review & editing.

Funding statement

The authors disclose receipt of the following financial or material support for the research, authorship, and/or publication of this article: C. McDaid, H. Leggett, and J. Lister report institutional funding for this project from Hull University Teaching Hospitals NHS Foundation Trust.

ICMJE COI statement

C. McDaid, H. Leggett, and J. Lister report institutional funding for this project from Hull University Teaching Hospitals NHS Foundation Trust. C. McDaid also reports research grants from NIHR HTA and i4i panels, which are unrelated to this study. C. McDaid is also Co-Director of the NIHR RSS Hub, a member of the NIHR PGfAR Panel B, the NIHR Predoctoral Fellowship Panel, Orthopaedic Research UK Scientific Advisory Committee, the Royal College of Surgeons of England Surgical Trials Centre group, and the British Orthopaedic Association Trials Network, the NIHR Journals Library Senior Journal Editor, and participates on the data safety monitoring board or advisory board of the NIHR TSC Op-Non STOP TSC and NIHR SCIENCE trial TSC.

Data sharing

The data that support the findings for this study are available to other researchers from the corresponding author upon reasonable request.

Ethical review statement

As this study was a review of existing patient-reported outcome measures and no additional data were collected; therefore, ethical approval and informed consent were not required.

Open access funding

Institutional funding for this project and its publication was received from Hull University Teaching Hospitals NHS Foundation Trust. The authors received a 10% discount as Prof. Sharma is a member of the British Orthopaedic Association.