

The Bone & Joint Journal  
22 BUCKINGHAM STREET  
LONDON WC2N 6ET, UK  
Telephone: +44 (0) 20 7782 0010  
Facsimile: +44 (0) 20 7782 0995  
Web: [www.bjj.boneandjoint.org.uk](http://www.bjj.boneandjoint.org.uk)

AUTHOR PROOFS

1. Please read the whole text carefully. **This is the last opportunity for change before publication**, and any changes should be typed directly into the text itself.

***Substantial changes should be discussed with Emily England, the Publishing Services Manager [e.England@boneandjoint.org.uk](mailto:e.England@boneandjoint.org.uk).***

2. Author queries (AQs) are shown as highlighted text with a comment box throughout the text. Please respond to the AQ directly, either by adding your response to the comment box or by making the correction / addition in the text itself.
3. Please ensure that all author information is correct, each author's details should include their degrees, job title, institution name and address. Please note this is the last opportunity to make changes of this nature.

***You must respond to ALL AQs, otherwise your paper will not progress to publication.***

4. Where indicated, please provide references which we have been unable to verify. Please email copies of the first and last pages of these, ensuring that they include full details, or provide web links to them online.  
**Please do NOT renumber the references. We will do this if needed.**
5. If you wish to replace a figure, please email it to Deborah Gray, and specify this within the galley proof
6. **Please make your corrections in the document, save it and email it back to Deborah Gray as soon as possible.**
7. We are happy to discuss any problems you may encounter with the galleys on the telephone.

**PLEASE NOTE THAT ALL CORRESPONDENCE MUST BE MARKED WITH YOUR PAPER LOG NUMBER.**

You will receive a free PDF of the final version of your paper soon after publication. The PDF should not be circulated beyond the authors on the paper and cannot be posted on any website other than *BJJ*.

I look forward to hearing from you regarding your galley proofs as soon as possible.

Best wishes,

Deborah Gray

Production Management Assistant

T: +44 (0) 20 7782 0010

F: +44 (0) 207 839 9777

E: [d.gray@boneandjoint.org.uk](mailto:d.gray@boneandjoint.org.uk)



The  
Bone & Joint  
Journal

70 years of orthopaedic excellence

Article ID

BJJ-2017-1349.R1

Article Title	The use of computer adaptive tests in outcome assessment following upper limb trauma
Article Type	Systematic review: trauma
JATS Article Type	
Article Subtitle	a systematic review
TOC Term	Systematic review: trauma
Subject Group	Trauma
Keywords	Trauma, Upper extremity, Computer adaptive testing, Systematic review
Associated Societies	
Author Name [1]	P. <b>Jayakumar</b>
Author Degree(s) [1]	MBBS BSc(Hons) MRCS (Eng) FMSEM DPhil
Author Job Title [1]	Orthopaedic Registrar and Honorary Research Fellow
Author Institution [1]	Nuffield Orthopaedic Centre
Author Address [1]	Headington, Oxford, <b>UK</b>
Author ID [1]	
Author Contribution [1]	Designing the study, Acquiring, analyzing, and interpreting the data, Drafting, revising, and approving the manuscript.
Author Name [2]	C. <b>Overbeek</b>
Author Degree(s) [2]	MD
Author Job Title [2]	Orthopaedic Doctor and Research Fellow
Author Institution [2]	Leiden University Medical Center
Author Address [2]	Leiden, <b>The Netherlands</b>
Author ID [2]	
Author Contribution [2]	Designing the study, Acquiring, analyzing, and interpreting the data, Revising and approving the manuscript.
Author Name [3]	A-M. <b>Vranceanu</b>
Author Degree(s) [3]	PhD
Author Job Title [3]	Associate Professor in Psychiatry
Author Institution [3]	Massachusetts General Hospital and Harvard Medical School
Author Address [3]	Massachusetts, <b>USA</b>
Author ID [3]	
Author Contribution [3]	Interpreting the data, Revising and approving the manuscript.
Author Name [4]	M. <b>Williams</b>
Author Degree(s) [4]	PhD MCSP
Author Job Title [4]	Programme Lead for Physiotherapy and Rehabilitation
Author Institution [4]	Oxford Brookes University
Author Address [4]	Headington, <b>UK</b>
Author ID [4]	
Author Contribution [4]	Designing the study, Acquiring, analyzing, and interpreting the data, Revising and approving the manuscript.
Author Name [5]	S. <b>Lamb</b>
Author Degree(s) [5]	MSc MA MCSP DPhil
Author Job Title [5]	Professor in Trauma Rehabilitation
Author Institution [5]	Nuffield Orthopaedic Centre
Author Address [5]	Headington, Oxford, <b>UK</b>
Author ID [5]	
Author Contribution [5]	Interpreting the data, Revising and approving the manuscript.
Author Name [6]	D. <b>Ring</b>
Author Degree(s) [6]	MD, PhD
Author Job Title [6]	Professor of Surgery and Associate Dean of Comprehensive Care, Department of Surgery and Perioperative Care
Author Institution [6]	Dell Medical School, The University of Texas at Austin
Author Address [6]	Austin, Texas, <b>USA</b>
Author ID [6]	
Author Contribution [6]	Interpreting the data, Revising and approving the manuscript.
Author Name [7]	S. <b>Gwilym</b>
Author Degree(s) [7]	DPhil FRCS (Orth)
Author Job Title [7]	Consultant Trauma and Orthopaedic Surgeon
Author Institution [7]	Nuffield Orthopaedic Centre
Author Address [7]	Headington, Oxford, <b>UK</b>
Author ID [7]	

<b>Author Contribution [7]</b>	Interpreting the data, Revising and approving the manuscript.
<b>Principal Institution</b>	University of Oxford and Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, Oxford, United Kingdom
<b>Correspondence</b>	Correspondence should be sent to P. Jayakumar; email: pjx007@gmail.com
<b>Abstract</b>	<p><b>Aims</b></p> <p>Outcome measures quantifying aspects of health in a precise, efficient, and user-friendly manner are in demand. Computer adaptive tests (CATs) may overcome the limitations of established fixed-length scales and be more adept at measuring outcomes in trauma. The primary objective of this review was to gain a comprehensive understanding of the psychometric properties of CATs compared with fixed-length scales in the assessment of outcome in patients who have suffered trauma of the upper limb. Study designs, outcome measures and methodological quality are defined, along with trends in investigation.</p> <p><b>Materials and Methods</b></p> <p>A search of multiple electronic databases was undertaken on 1 January 2017 with terms related to “CATs”, “orthopaedics”, “trauma”, and “anatomical regions”. Studies involving adults suffering trauma to the upper limb and undergoing any intervention were eligible. Those involving the measurement of outcome with any CATs were included. Identification, screening, and eligibility were undertaken, followed by the extraction of data and quality assessment using the Consensus-Based Standards for the Selection of Health Measurement Instruments (COSMIN) criteria. The review is reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) criteria and registered (PROSPERO: CRD42016053886).</p> <p><b>Results</b></p> <p>A total of 31 studies reported trauma conditions alone or in combination with non-traumatic conditions using CATs. Most were cross-sectional with varying level of evidence, number of patients, type of study, range of conditions and methodological quality. CATs correlated well with fixed-length scales and had minimal or no floor-ceiling effects. They required significantly fewer questions and/or less time for completion. Patient-Reported Outcomes Measurement Information System (PROMIS) CATs were the most frequently used, and the use of CATs is increasing.</p> <p><b>Conclusion</b></p> <p>Early studies show valid and reliable outcome measurement with CATs performing as well as, if not better than, established fixed-length scales. Superior properties such as floor-ceiling effects and ease of use support their use in the assessment of outcome after trauma. As CATs are being increasingly used in patient outcomes research, further psychometric evaluation, especially involving longitudinal studies and groups of patients with specific injuries are required to inform clinical practice using these contemporary measures.</p>

	<p>Key messages:</p> <ul style="list-style-type: none"> <li>-- CATs demonstrate valid and reliable outcome measurement and perform as well, if not better, than established fixed length PROMs</li> <li>- Early studies suggest the validity and reliability of CATs in assessing patient outcomes following orthopaedic trauma</li> <li>- Properties such as improved floor-ceiling effects and ease of use further support their utilisation in trauma settings</li> <li>- Further psychometric evaluation especially involving longitudinal studies and populations with specific trauma conditions are required</li> </ul> <p>Cite this article: <i>Bone Joint J</i> 2018;100-B:??-??.</p>
<b>NLM Journal ID</b>	Bone Joint J
<b>Print ISSN</b>	2049-4394
<b>Article Year</b>	2018
<b>Online ISSN</b>	2049-4408
<b>Issue No</b>	6
<b>Volume No</b>	100-B
<b>Page Range</b>	??-??
<b>Price</b>	\$2.00
<b>Copyright Statement</b>	©2018 The British Editorial Society of Bone & Joint Surgery
<b>Funding Statement</b>	<p>No benefits in any form have been received or will be received from a commercial party related directly or indirectly to the subject of this article.</p> <p>This article was primary edited by J. Scott.</p>
<b>ICMJE COI Statement</b>	None declared
<b>Date Received</b>	
<b>Date Accepted</b>	
<b>Number of figures</b>	3
<b>DOI</b>	10.1302/0301-620X.100B6.BJJ-2017-1349.R1

Patient-reported outcome measures (PROMs) have been developed to capture and evaluate the subjective impact of conditions and treatments.[[2-5]] In orthopaedic surgery, several systems of measurement have been designed to capture aspects of physical, psychological, and social health using a fixed set of questions.[[5]] These ‘fixed-length scales’ have commonly been developed using ‘classic test theory’, where a response is required for most, if not all, questions to achieve a valid score.[[2,5]]. This form of PROM is often limited in scope and usually involves a trade-off between the range of questions and the precision of the scores.[[6]] Administration of such measurements, especially in batteries of tests incorporating many questions, may burden both patients and investigators. Although shortened versions of these scales have been developed, they may provide less accurate data if the number of questions is not reduced carefully.[[7-9]]

Computer adaptive tests (CATs) have been developed to overcome these limitations. [[6]] Introduced in the 1970s in education and aptitude testing, they have more recently been applied to measure outcomes in health care.[[10,11]]. They are based on ‘item response theory’ (IRT), a mathematical model enabling the formation of adaptive computerized algorithms designed to optimize the administration of questionnaires.[[10-13]] Only relevant questions are used based on a patient’s previous response and their estimated level within a specific health domain.[[10,11,13]] For instance, if the patient states they can run ten miles easily, they will not be asked if they can walk 100 yards.

CATs often use large cohorts of patients to form extensive, calibrated banks of questions. This extends the range of assessment, providing broader and more detailed cover. The resultant tests are designed to provide the tailored, precise, and efficient measurement of outcome.[[14-17]]. The wide range of questions also aims to minimize floor and ceiling effects.[[18-20]] Fixed-length scale measures may be limited in this respect, as the quality of measurement at the extremes is often poor.[[9,11,14]] Types of CAT include generic, modular measures targeted at specific health domains, such as the Patient Reported Outcome Measurement Information System (PROMIS) CATs, and region-specific measures, such as the Shoulder Functional Status CAT (SFS-CAT) and therapy-specific measures, such as the Activity Measure for Post-Acute Care CAT (AM-PAC-CAT) for rehabilitation. The PROMIS measures, produced by the National Institute of Health, were originally developed with chronic conditions in mind.[[6]] However, based on their measurement capabilities, they may also be used to record outcomes in patients who have suffered trauma with increased speed and accuracy compared with fixed-length scales.[[15-17]]. Trauma affects a wide range of patients and is often associated with significant shifts in health status. For instance, low-energy fragility fractures in the elderly may show a different disability profile compared with high-energy fractures in the younger patient. The positive effects of the stable fixation of fractures, healing and recovery may be countered by pain, physical events (e.g. complications), psychological issues (e.g. ineffective coping) and social

constraints (e.g. secondary gain). CATs, such as PROMIS, provide a modular suite of measures covering all the major domains of health and disability. For instance, if investigators wish to assess the level of physical functioning, depression and pain related to a condition, the relevant CATs can be selected and administered.

The primary aim of this review was to gain a comprehensive understanding of the psychometric properties of CATs compared with fixed-length scales in trauma involving the upper limb. Secondly, we define the study designs, diagnoses, research settings and outcome measures investigated, including the most popular CATs being used. The assessment of quality was performed using the Consensus-Based Standards for the Selection of Health Measurement Instruments (COSMIN) criteria and rating system.[[21-23]] Finally, we examined the trend in the investigation of CATs in orthopaedics over the last decade.

## **Materials and Methods**

We performed a broad search strategy using the PubMed, OVIDSP, CINAHL, PsycInfo, and Cochrane Library databases on 1 January 2017. A lack of MeSH terms for CATs in the assessment of PROMs and a wide variation in labelling led to the use of a broad strategy. Search terms were developed with a librarian and validated by a multidisciplinary team. Terms related to ‘CATs’ (computer\* adapt\* test\* OR “item response theory (IRT)” OR IRT OR PROMIS), ‘orthopaedics’, ‘trauma’, and ‘anatomical regions’ (trauma\* OR orthop\* OR

musculoskeletal OR injury OR limb\* OR extremity OR arm\* OR leg\* OR shoulder\* OR elbow\* OR wrist\* OR hand\* OR spine OR spinal OR pelvis OR pelvic OR hip\* OR knee\* OR ankle\* OR foot) [ti,ab] were deployed using the operator “AND”. No restrictions were set in the search fields and terms were identified in the title and/or abstract without limits. The review adheres to Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) criteria,[[24]] and the protocol is registered on the PROSPERO system (Reg: CRD42016053886).

### Study selection

The selection of studies involved adults having any trauma involving the upper limb, managed conservatively or surgically, at any stage along the care pathway (e.g. surgical episode, rehabilitation). Outcome measurement with any CAT was included and only original research in peer-reviewed journals in the English language was considered eligible. Non-clinical, non-orthopaedic, non-CAT related studies and investigations involving fixed-length scales and short-form versions of CATs were excluded. The lead author (PJ) performed the search and two investigators (PJ, MAW) screened the titles and abstracts using a step-wise exclusion process (Fig. 1). Three investigators (PJ, SG, MAW) independently reviewed the texts and bibliographies for additional publications. Disagreements were resolved by consensus meetings coordinated by the lead author (PJ). Authors of original articles were contacted if the studies did not clearly specify the condition investigated or reported the outcomes insufficiently.



[[Fig 1]]

[[FigCap]]Flowchart based on the PRISMA framework showing the phased process of screening, identification, eligibility and inclusion in the final study set. Study type I: biomedical, i.e. basic science; biomedical engineering; computer/mathematic modelling; non-computer adaptive test (CAT)-based technology assessment; performance testing; imaging study. Study type II: review; technical report; case report; commentary; erratum. Study type III: Rasch analysis/item response theory (IRT) application in orthopaedic area but non-CAT related. Study type IV: Rasch analysis/IRT application in non-orthopaedic area; statistical study only. Study type V: study involving CATs but non-orthopaedic clinical focus. Study type VI: study involving orthopaedic clinical focus but non-CAT related. Study type VII: study involving neither orthopaedic or CAT-related clinical focus.

### Data extraction and data synthesis

Two investigators (PJ, MAW) independently extracted data from the eligible studies onto an electronic database (Microsoft Excel; Microsoft, Redmond, Washington) including level of evidence, number of patients, type of study, region and conditions, outcome measures and therapeutic setting. Further documentation of psychometric properties including validity, reliability, responsiveness, interpretability and feasibility characteristics (i.e. the number of questions or time taken to complete a test) in comparison with fixed-length scales was conducted. Formal dates of publication were recorded to map the distribution of the studies over time.

Classification of the type of study was performed once the total study set was assessed for common themes. This included studies involving

‘development’, ‘testing’, ‘outcome assessment’ and combinations thereof.

‘Development’ involved studies using various methods of developing outcome measures to construct a CAT. Studies classified by ‘testing’ involved analysis of the psychometric properties of CATs, such as validity, reliability and responsiveness. Those termed as ‘outcome assessment’ studies entailed the use of CATs to measure outcomes for a condition or intervention either in combination with fixed-length scales or alone.

### **Quality assessment**

Quality assessment was conducted using the COSMIN criteria and four-point rating system.[[21,22]] The checklist provides a standardized assessment of methodological quality, design requirements and preferred statistical analysis of measurement properties of PROM instruments, including those involving IRT models. The rating system uses a series of questions for scoring the relevant measurement properties (i.e. IRT methodology, validity, reliability and responsiveness) within each study in a modular fashion. Ratings of ‘poor’, ‘fair’, ‘good’ or ‘excellent’ are provided with the lowest rating of any question in an assessment category taken as the final score, i.e. ‘worst score counts’.

Generalizability (i.e. characteristics of the study population and sampling procedure) and interpretability (e.g. normal scores, floor-ceiling effects) data were extracted but not scored.

## Results

The search provided a total of 3777 citations and after exclusion of duplicate studies, non-English-language studies, and non-human studies, 1763 remained. Of these, 1627 were excluded following screening of the titles and abstracts, leaving 136 studies for full text review. Of these, 31 (23%) either solely involved traumatic conditions ( $n = 6/136$ ; 4%) or involved a combination of traumatic and non-traumatic conditions ( $n = 25/136$ ; 18%) (Fig. 1)(Table I).

**Table I.** Characteristics of studies involving computer adaptive tests (CATs) and trauma affecting the upper limb.

Author (yr)	Study design (level of evidence)[[51]]	Patients (n)	CAT study type	Upper extremity region	Upper extremity trauma condition(s)	Trauma conditions alone or in combination	CAT	Fixed-length scale and other outcome metrics	Management therapy setting (country)
Beckmann et al[[29]] (2015)	Observational, cross-sectional; outcomes research (level 2c)	187	Testing	Shoulder	Rotator cuff tear	C	PROMIS PF CAT	ASES; SST	Outpatient (USA)
Cook et al[[32]] (2005)	Prospective cohort study (level 2b)	400	Combination: development & testing	Shoulder	N/A	C	Shoulder CAT	FLEX-SF; ASES; SF-12	N/A (USA)
Döring et al[[33]] (2014)	Observational, cross-sectional; outcomes research (level 2c)	84	Testing	Multi-region	Hand, wrist fractures; mallet finger; soft-tissue injury; animal bite	C	PROMIS UE PF CAT; PROMIS PF Mobility CAT; PROMIS PI CAT	QuickDASH; PSEQ-2; PHQ-2	Outpatient (USA)
Haley et al[[34]] (2009)	Observational, cross-sectional; outcomes research (level 2c)	6 882	Development	Multi-region	N/A	C	DA-CAT1 (AM-PAC CAT); DA-CAT2	N/A	Outpatient clinic (US)

Hart et al[[35]] (2006)	Secondary analysis of prospective cohort study (level 2b)	400	Development	Shoulder	N/A	C	SFS CAT	(FLEX-SF items used for development only)	Outpatient clinic (US)
Hart et al[[36]] (2010)	Secondary analysis of prospective cohort study (level 2b)	30 9 87	Development	Shoulder	Rotator cuff tear; sprains and strains; ACJ injury; clavicular, scapular,, humeral fracture	C	Shoulder CAT; CAT-FS	(GROC only for assessing change amongst individual patients)	Outpatient clinic (US)
Hays et al[[37]] (2013)	Observational, cross-sectional; outcomes research (level 2c)	21 7 73	Development	Multi-region	N/A	C	PROMIS PF CAT	Nil	N/A (USA)
Hung et al[[16]] (2011)	Observational, cross-sectional; outcome research (level 2c)	865	Development	Multi-region	N/A	C	PROMIS PF CAT	Nil	Outpatient (USA)
Hung et al[[18]] (2014)	Observational, cross-sectional; outcomes research (level 2c)	153	Combination: testing & outcomes assessment	Multi-region	N/A	Trauma	PROMIS PF CAT	sMFA	Outpatient (USA)
Janssen et al[[38]] (2015)	Observational, cross-sectional;	139	Outcome assessment	Multi-region	Fracture shoulder, elbow, forearm, wrist, hand; finger	C	PROMIS PB CAT; PROMIS PI	QuickDASH; pain rating scale	Outpatient (USA)

	outcomes research (level 2c)				amputation; MCPJ dislocation; soft-tissue injury hand		CAT; PROMIS PF CAT; PROMIS Depression CAT	(11-point ordinal scale)	
Jayakumar et al[[39]] (2014)	Observational, cross-sectional; outcomes research (level 2c)	98	Outcomes assessment	Multi-region	Elbow, hand, humeral, forearm, wrist fractures; soft-tissue injury hand; animal bite	C	PROMIS PF CAT; PROMIS PI CAT; PROMIS Depression CAT	PACES	Outpatient (USA)
Jette et al[[40]] (2007)	Prospective cohort study; outcome research (level 2c)	1815	Development	Multi-region	N/A	C	AM-PAC-CAT	N/A	Outpatient clinic (US)
Kortlever et al[[41]] (2015)	Observational, cross-sectional; outcome research (level 2c)	138	Combination: testing & outcomes assessment	Multi-region	N/A	C	PROMIS PI CAT	PCS; PIPS; PSEQ; QuickDASH	Outpatient (USA)
Lehman et al[[42]] (2011)	Observational, cross-sectional; outcome research (level 2c)	312	Development	Multi-region	Upper extremity fractures	C	ICF-based UE CAT	N/A	Outpatient (USA)

Mellema et al[[43]] (2015)	Prospective cohort study (level 1b)	136	Outcome assessment	Wrist and Hand	Wrist fracture; hand fracture; sprain, rupture, dislocation; amputation, crush, laceration	C	PROMIS PI CAT	Pain VAS; patient Satisfaction (11-point ordinal scale); patient-physician communication	Outpatient (USA)
Menendez et al[[44]] (2013)	Observational, cross-sectional; outcome research (level 2c)	213	Combination: testing & outcomes assessment	Multi-region	Fracture shoulder, elbow, wrist, hand; sprain, dislocation, or mallet finger; amputation, crush, or laceration	C	PROMIS PI CAT; PROMIS Depression CAT	QuickDASH; PSEQ	Outpatient (USA)
Menendez et al[[45]] (2015)	Observational, cross-sectional; outcomes research (level 2c)	200	Outcome assessment	Multi-region	Fracture elbow, wrist, hand; sprain, dislocation, or mallet finger; amputation, crush, or laceration	C	PROMIS UE PF CAT; PROMIS PI CAT	NVS health literacy assessment tool	Outpatient (USA)
Menendez et al[[46]] (2015)	Observational, cross-sectional; outcomes research (level 2c)	122	Combination: testing & outcomes assessment	Multi-region	Sprain, dislocation, mallet finger; hand, wrist, elbow fracture; amputation/crush injury	C	PROMIS PI CAT; PROMIS Depression CAT; PROMIS UE PF CAT	N/A	Outpatient (USA)
Menendez et al[[47]] (2015)	Observational, cross-sectional; outcomes research (level 2c)	112	Outcome assessment	Wrist and Hand	Hand conditions – traumatic and non-traumatic	C	PROMIS PI CAT; PROMIS UE PF CAT; PROMIS Depression CAT	Satisfaction rating – CGCAHPS; CARE; NVS health literacy test; sociodemographic survey; waiting	Outpatient (USA)

								time in office; duration of visit; time from booking until appointment	
Moradi et al[[48]] (2015)	Observational, cross-sectional (level 3b)	155	Outcomes assessment	Wrist and Hand	Hand fracture; distal radius fracture; sprain, rupture, or dislocation; elbow fracture; amputation, crush, or laceration	C	PROMIS Depression CAT	Pain and numbness on hand diagrams; PCS; SHAI-5	Outpatient (USA)
Morgan et al[[26]] (2015)	Prospective cohort study (level 2b)	47	Combination: testing + outcomes assessment	Shoulder	Proximal humeral fracture	Trauma	PROMIS PF CAT	CMS; DASH; sMFA; ROM; shoulder strength	Operative nonoperative outpatient clinic (USA)
Okike et al[[27]] (2015)	Observational, cross-sectional (level 3b)	207	Outcome assessment	Shoulder	Proximal humeral fracture	Trauma	PROMIS PF CAT	CMS; DASH; sMFA; nonunion; malunion; humeral head AVN; complications (screw perforation, loss of fixation, infection, neurovascular injury, secondary surgery)	Operative (locking plate fixation); nonoperative outpatient clinic (USA)



Overbeek et al[[49]] (2014)	Observational, cross-sectional study; outcome research (level 2c)	93	Combination: testing + outcomes assessment	Multi-region	Hand, wrist fractures; mallet finger; soft-tissue injury; animal bite	C	PROMIS PF CAT; PROMIS Depression CAT; PROMIS PI CAT	QuickDASH; NRS (11-point ordinal scale)	Outpatient (USA)
Peters et al[[50]] (2016)	Observational, cross-sectional; outcome research (level 2c)	111	Combination: testing + outcomes assessment	Multi-region	Sprain, dislocation or mallet finger; hand fracture; wrist fracture; amputation, crush or laceration; carpal tunnel or cubital tunnel syndrome; OA; trigger finger; non-specific arm pain; elbow fracture; De Quervain's tendinopathy; Dupuytren's disease; other	C	PROMIS Sleep disturbance CAT; PROMIS UE PF CAT; PROMIS PI CAT; PROMIS Depression CAT	N/A	Outpatient (USA)
Quispe et al[[1]] (2015)	Case series (level 4)	3	Outcome assessment	Shoulder	Sternoclavicular joint dislocations	Trauma	PROMIS PF-Mobility CAT	DASH; restoration of complete bony alignment and stability (chest x-ray)	Operative transarticular plating (US)
Sindhu et al[[51]] (2012)	Retrospective cohort study,	3362	Outcome assessment	Shoulder	Fracture of the clavicle, shoulder,	C	Shoulder CAT	NRS (11-point ordinal); FABQ-PA	Outpatient clinic (US)

	longitudinal (level 2b)				proximal humerus; sprains and strains				
Sindhu et al[[52]] (2013)	Secondary analysis of prospective cohort study (level 2b)	3 767	Combination: development & testing	Shoulder	Tendon/synovial rupture; traumatic wounds or amputation; contusions; crush injury; fracture / dislocation repair or reconstruction; proximal humerus, greater tuberosity, lesser tuberosity, humeral head, clavicular fracture; ACJ disruption	C	Shoulder CAT	NRS (0 to 10 scale)	Outpatient clinic (US)
Stuart et al[[53]] (2015)	Prospective cohort study (level 1b)	55	Combination: testing & outcomes assessment	Multi-region	Fractured clavicle, distal humerus, radius, proximal humerus, intra-articular elbow	Trauma	PROMIS PF CAT	N/A	Outpatient (USA)
Tyser et al[[54]] (2014)	Observational, cross-sectional; outcomes research (level 2c)	134	Testing	Multi-region	Nos	C	PROMIS PF CAT	DASH	Outpatient (USA)
Van Leeuwen et al[[55]] (2016)	Observational, cross-sectional; outcome	124	Combination: testing & outcomes assessment	Multi-region	Upper extremity injury; lower extremity injury incl. fractures	Trauma	PROMIS PF CAT; PROMIS PI CAT	IEQ; PHQ-2; PSEQ-2; PCS-4	Outpatient (USA)

	research (level 2c)								
Wang et al[[25]] (2010)	Prospective cohort study; outcomes research (level 2c)	30 9 87	Combination: testing & outcomes assessment	Shoulder	Tendon/synovial rupture; traumatic wounds or amputation; contusions; crush injury; fracture / dislocation repair or reconstruction; proximal humerus, greater tuberosity, lesser tuberosity, humeral head, clavicle fracture; ACJ disruption	C	Shoulder CAT	N/A	Outpatient clinic (US)

[[TblNote]]C, combination of traumatic and non-traumatic conditions; PROMIS, Patient Reported Outcome Measurement Instrumentation System; PF, physical function; ASES, American Shoulder Elbow Society score; SST, Simple Shoulder Test score; N/A, not applicable or not available; FLEX-SF, flexi-level scale of shoulder function; SF-12, 12-Item Short-Form Health Survey; UE, upper extremity; PI, pain interference; (Quick)DASH, (Quick) Disabilities of the Arm, Shoulder and Hand score; PSEQ, Pain Self Efficacy Questionnaire; PHQ, Patient Health Questionnaire; DA, daily activity; AM-PAC, activity measure for post-acute care; SFS, shoulder functional status; FS, functional status; GROG, Global Rating of Change; sMFA, Short Musculoskeletal Functional Assessment; MCPJ, metacarpophalangeal joint; PB, pain behaviour; PACES, Physical Activity Enjoyment Scale; PCS, Pain Catastrophizing Questionnaire; PIPS, Psychological Inflexibility in Pain Scale; ICF, International Classification of Functioning, Disability and Health; VAS, visual analogue scale; NVS, newest vital sign; CGCAHPS, Clinician and Group-Consumer Assessment of Healthcare Providers and Systems survey; CARE, Consultation and Relational Empathy measure; SHAI, Short Health Anxiety Inventory; CMS, Constant–Murley score; ROM, range of movement; AVN, avascular necrosis; NRS, Numerical Rating scale; FABQ-PA, Fear Avoidance Beliefs Questionnaire Physical Activities scale; ACJ, acromioclavicular joint; IEQ, Injustice Experience Questionnaire

Most studies were cross-sectional investigations (20/31; 65%) and level of evidence was variable (Level I, 2/31 (6.5%); Level II, 26/31 (84%); Level III, 2/31(6.5%); Level IV, 1/31 (3%)) (Table I). The number of patients ranged from three to 30 987.[[1,25]] The largest category of study involved a combination of ‘testing and outcomes assessment’ (10/31; 33%) (Table I) (Fig. 1). Most studies involved conditions in several regions (18/31; 58%), followed by those affecting the shoulder (10/31; 32%) and wrist and hand (3/31; 10%). Injuries ranged from fractures, dislocations and subluxations to soft-tissue trauma, with a few focusing on specific injuries such as proximal humeral fractures[[26,27]] and dislocations of the sternoclavicular joint[[1]] (Table I). The most frequently used CATs were the PROMIS PF CAT, PROMIS Pain Interference CAT and PROMIS Depression CAT (Fig. 2). All studies were performed in the USA with most data being obtained in outpatient clinics.

[[Fig 2]]

[[FigCap]]Graph showing the frequency of computer adaptive tests (CATs) used in the final inclusion. PROMIS, Patient Reported Outcome Measurement Instrumentation System; PF, physical function; UE, upper extremity; SFS, shoulder functional status; DA, daily activity; ICF, International Classification of Functioning, Disability and Health.

### Comparison of CATs with fixed-length scales

Nine studies (29%) involved the use of CATs and fixed-length scale measures and included the evaluation of one or more psychometric properties (Table II). Moderate to strong correlations were found between CATs (e.g. PROMIS PF

CAT) and fixed-length scales (e.g. Disabilities of the Arm Shoulder and Hand (DASH) score). CATs also consistently showed minimal or no floor-ceiling effects, with significantly shorter times to complete the test or fewer items for completion compared with fixed-length scales (Table II).

**Table II.** Comparison of psychometric properties in studies involving computer adaptive tests (CATs) and fixed-length scales

	Psychometric properties (validity, reliability, responsiveness)	Interpretability (floor-ceiling effects)	Feasibility/ease of use
Beckmann et al[[29]] (2015)	PROMIS PF CAT, ASES, and SST demonstrate moderate correlation. Item reliability was excellent for all instruments. Person reliability of PROMIS PF CAT was superior (0.93, excellent) to SST (0.71, moderate) and ASES (0.48, fair).	Ceiling effects were similar among all instruments (PF, 0.53%; SST, 6.1%; ASES, 2.3%). Floor effects found in 21% for SST, 3.2% for PF CAT, and 2.3% for ASES.	PROMIS PF CAT required fewer questions than ASES or SST (PROMIS PF CAT, 4.3; ASES, 11; SST, 12) for test completion.
Cook et al[[32]] (2005)	N/A	N/A	Shoulder CAT required only a few items before the SEM-based stopping rule was reached (6% were administered < 3 items, 11% < 4 items and 73% < 10 items).
Döring et al[[33]] (2014)	PROMIS UE PF CAT & QuickDASH demonstrated strong correlation ( $\beta = -.81$ ; $p < 0.001$ ) & both had moderate correlation with PROMIS PF CAT & PROMIS PF Mobility CAT. PROMIS PI CAT was the strongest correlate of QuickDASH ( $\beta = -0.66$ ; $p < 0.001$ ) & PROMIS UE PF CAT ( $\beta = 2.0$ ; $p < 0.001$ ).	PROMIS UE PF CAT showed no floor (0%) or ceiling effects (0%). QuickDASH showed no ceiling effects and one patient scored 0.	PROMIS PF CAT demonstrated significantly shorter completion times than QuickDASH (mean (SD, range): QuickDASH, 116 (47, 42 to 281) vs PROMIS UE PF CAT, 70 secs (47, 25 to 300), $p < 0.001$ ).
Hung et al[[18]] (2014)	PROMIS PF CAT and sMFA showed extremely high item reliability (Cronbach's $\alpha = 0.98$ ).	Neither instrument showed any floor effect. sMFA revealed 14.4% ceiling effect and PROMIS PF CAT had no appreciable ceiling effect.	PROMIS PF CAT administration time was 44 secs vs 599 secs for sMFA ( $p < 0.05$ ).
Kortlever et al[[41]] (2015)	Substantial correlation between PCS (0.74), PIPS (0.84), PROMIS PI CAT (0.83), PSEQ (-0.86) & underlying trait. Interquestionnaire correlation was large to substantial: highest, PROMIS PI CAT with PSEQ ( $\rho = -0.84$ , $p < 0.001$ ); lowest, PROMIS PI CAT with PCS ( $\rho = 0.67$ , $p < 0.001$ ). Internal consistency was high. PROMIS PI CAT had strongest correlation with QuickDASH (B 0.63; SE 0.10, $p < 0.001$ ).	N/A	PROMIS PI CAT was the quickest to complete (median (IQR) 30 secs (22 to 44); PCS, 91 secs (66 to 120), $p < 0.001$ ; PIPS, 105 secs (82 to 141), $p < 0.001$ ; PSEQ, 78 secs (54 to 101), $p < 0.001$ ).
Menendez et al[[44]] (2013)	Large correlation between QuickDASH & PROMIS PI CAT ( $r = 0.74$ ; $p < 0.001$ ) and between PROMIS PI CAT & PSEQ ( $r = -0.72$ , $p < 0.001$ ). PROMIS Depression CAT had medium correlation with QuickDASH ( $r = 0.37$ ; $p < 0.001$ ) & PROMIS PI CAT ( $r = 0.40$ ; $p < 0.001$ ). 51% variability in QuickDASH is explained by PROMIS PI alone.	N/A	N/A
Morgan et al[[26]] (2015)	PROMIS PF CAT correlated significantly (moderate to high) with all other physical function outcome measures (i.e. DASH, sMFA, CMS).	N/A	Patients responded to fewer PROMIS PF CAT items (mean 4 items) than other scores (mean 6 for CSS, 30 for DASH, and 46 for sMFA). Time to completion of the PROMIS PF CAT was 98 secs (mean 98, which was significantly less than DASH (median 482 secs, $p < 0.001$ ).
Overbeek et al[[49]] (2014)	QuickDASH moderately correlated with PROMIS PF CAT ( $r = -0.55$ , $p < 0.001$ ). PROMIS PF CAT and QuickDASH correlated strongly with PROMIS depression ( $r = -0.35$ , $p < 0.001$ ; $r = 0.34$ , $p < 0.001$ ) and PROMIS PI CAT ( $r = -0.51$ , $p < 0.001$ ; $r = 0.74$ , $p < 0.001$ ), respectively. Factors accounting for variability in PROMIS scores were comparable to those for Quick DASH.	N/A	N/A
Tyser et al[[54]] (2014)	DASH and PROMIS PF CAT had a strong correlation ( $r = 0.726$ ). Item and person reliability were 0.97 and 0.94 respectively for DASH, and 0.99 and 0.96 for PROMIS PF CAT. DASH and PF CAT had 5% and 5% of unexplained variance respectively.	DASH exhibited 5% ceiling effect and 1% floor effect whereas PF CAT had no ceiling or floor effects.	DASH, mean time to completion was 262 secs; PROMIS PF CAT mean 57 secs.

PROMIS, Patient Reported Outcome Measurement Instrumentation System; PF, physical function; ASES, American Shoulder Elbow Society score; SST, Simple Shoulder Test score; N/A, not applicable or not available; SEM, ...; UE, upper extremity; (Quick)DASH,

(Quick)Disabilities of the Arm, Shoulder and Hand score; sMFA, Short Musculoskeletal Functional Assessment; PCS, Pain Catastrophizing Questionnaire; PIPS, Psychological Inflexibility in Pain Scale; PI, pain interference; PSEQ, Pain Self Efficacy Questionnaire; IQR, interquartile range; CMS, Constant–Murley score; CSS, Constant Shoulder Score

### Methodological quality

Quality assessment showed an adequate definition of the characteristics of the patients and sampling overall, but descriptions of conditions and diseases were lacking in 15 studies (43%) where contact with the corresponding authors was required for clarification. Despite this contact, nine studies (29%) remained unspecified as the investigators were unable to recall the diagnoses. Sample sizes within studies were adequate with few methodological flaws and sound statistical evaluation; however, reporting and the description of how missing items and measurements were handled over time was poor.

Studies involving IRT-based analysis rated ‘good’ or ‘excellent’, although descriptions of the model, computer software package, method of estimation and assumptions were very variable. One or more types of reliability (i.e. internal consistency, test-retest, intra or interobserver reliability, and measurement error) were assessed in 15 studies (48%), and these were very variable in quality, mostly rating ‘fair’ or ‘good’ (Table III). Validity was assessed in some form in all studies and mostly rated ‘good’ or ‘excellent’.

Greater variability in ratings was observed in the assessment of structural validity and hypothesis-testing compared with content validity. Criterion validity was not assessed due to a lack of a ‘gold standard’. Cross-cultural validity was not within the scope of our study. The assessment of interpretability showed variable accounts of the distribution and changes of scores, with few authors undertaking analysis of floor-ceiling effects and fewer



still calculating the minimal clinically important difference (MCID) or minimal detectable change (MDC). The heterogeneity of the study in terms of design, diagnoses, outcome measures and statistical analysis meant that pooling was not considered feasible or appropriate. Best-evidence synthesis to summarize the strength of evidence for specific CATs could not be performed for the same reason. Research involving the use of CATs in orthopaedic surgery during the last decade appears to be increasing year-on-year (Fig. 3).

**Table III. Quality assessment using the COSMIN criteria and four-point checklist**

Author (Year)	IRT Box	Reliability			Validity				Responsiveness	
		A. Internal consistency	B. Reliability	C. Measurement error	D. Content validity	E. Structural validity	F. Hypothesis testing	H. Criterion validity	I. Responsiveness	J. Interpretability
Beckmann et al[[29]] (2015)*	Good	Fair	N/A	N/A	Good	Fair	Good	N/A	N/A	SD; FC
Cook et al[[32]] (2005)*	Good	Fair	Fair	Fair	Excellent	Poor	Fair	N/A	Fair	SD; SC; MCID
Döring et al[[33]] (2014)*	N/A	N/A	N/A	N/A	Good	N/A	Good	N/A	N/A	SD; SC; FC
Haley et al[[34]] (2009)*	Excellent	Good	N/A	N/A	Excellent	Good	Fair	N/A	N/A	SD; SC; FC
Hart et al[[35]] (2006)*	Excellent	Excellent	N/A	N/A	Excellent	Excellent	Fair	N/A	Fair	SD; SC; MDC
Hart et al[[36]] (2010)	N/A	N/A	N/A	Good	Good	N/A	N/A	N/A	Excellent	SD; SDC; FC; MDC; MCID
Hays et al[[37]] (2013)*	Excellent	Fair	N/A	N/A	Good	Fair	N/A	N/A	N/A	SD; SC; FC
Hung et al[[16]] (2011)*	Good	Excellent	N/A	N/A	Excellent	Excellent	N/A	N/A	N/A	SD; SC; FC
Hung et al[[18]] (2014)*	Good	Good	N/A	Good	Good	Good	N/A	N/A	N/A	SD; FC
Janssen et al[[38]] (2015)*	N/A	N/A	N/A	N/A	Good	N/A	Good	N/A	N/A	SD; SC
Jayakumar et al[[39]] (2014)	N/A	N/A	N/A	N/A	Good	N/A	Good	N/A	N/A	SD; SC
Jette et al[[40]] (2007)*	Good	N/A	N/A	Good	Excellent	Good	N/A	N/A	Good	SD; SC; FC; MDC
Kortlever et al[[41]] (2015)*	N/A	Good	N/A	N/A	Excellent	Good	Good	N/A	N/A	SC; FC
Lehman et al[[42]] (2011)	Excellent	Excellent	N/A	N/A	Excellent	Excellent	N/A	N/A	N/A	SD; FC
Mellema et al[[43]] (2015)	N/A	N/A	N/A	N/A	Good	N/A	Good	N/A	N/A	SD; SC
Menendez et al[[44]] (2013)	N/A	N/A	N/A	N/A	Good	N/A	Excellent	N/A	N/A	SC
Menendez et al[[45]] (2015)	N/A	N/A	N/A	N/A	Good	N/A	Good	N/A	N/A	SD
Menendez et al[[46]] (2015)	N/A	N/A	N/A	N/A	Good	N/A	Fair	N/A	N/A	SC
Menendez et al[[47]] (2015)	N/A	N/A	N/A	N/A	Good	N/A	Good	N/A	N/A	SD; SC
Moradi et al[[48]] (2015)	N/A	N/A	N/A	N/A	Excellent	N/A	Good	N/A	N/A	SC
Morgan et al[[26]] (2015)	N/A	N/A	N/A	N/A	Good	N/A	Good	N/A	N/A	SD; SC; FC
Okike et al[[27]] (2015)	N/A	N/A	N/A	N/A	Poor	N/A	N/A	N/A	N/A	SC
Overbeek et al[[49]] (2014)*	N/A	N/A	N/A	N/A	Good	N/A	Good	N/A	N/A	SD; SC
Quispe et al[[1]] (2015)	N/A	N/A	N/A	N/A	Poor	N/A	N/A	N/A	N/A	SD; SC
Sindhu et al[[51]] (2012)	N/A	N/A	N/A	N/A	Good	N/A	Good	N/A	N/A	SD
Sindhu et al[[52]] (2013)*	Good	Good	Good	Good	Good	Poor	Good	N/A	Poor	SD; SC; MDC
Sindhu et al[[52]] (2013)*	Good	N/A	N/A	N/A	Excellent	N/A	Fair	N/A	N/A	SD; SC
Stuart et al[[53]] (2015)	N/A	N/A	Good	N/A	Good	N/A	N/A	N/A	N/A	SD; SC
Tyser et al[[54]] (2014)*	Good	Good	N/A	N/A	Good	Good	Fair	N/A	N/A	SD; SC; FC
van Leeuwen et al[[55]] (2016)	N/A	N/A	N/A	N/A	Good	N/A	Good	N/A	N/A	SC
Wang et al[[25]] (2010)*	Excellent	Excellent	N/A	N/A	Excellent	N/A	N/A	N/A	N/A	SD; SC; MDC; MCID

\*Authors contacted for clarification of upper extremity conditions in their study cohort

N/A, not applicable or not available; SD, score distribution; FC, floor-ceiling effect; SC, score change; MCID, minimal clinically important difference; MDC, minimal detectable change

[[Fig 3]]

[[FigCap]]Graph showing the number of full text publications in peer-reviewed journals involving computer adaptive tests in orthopaedics by year of publication since 2005.

## **Discussion**

The measurement of PROMs is integral to modern health care. There is an increasing demand for measures that are valid, reliable, responsive and easy to use in specific groups of patients. CATs appear to perform precise and efficient outcome measurement following injuries to the upper limb. The correlation with established fixed-length scales, better floor-ceiling effects and test completion times provide support for their use in patients who sustain upper extremity trauma. The PROMIS measures are by far the most frequently used CATs, providing an easy reference to normal values in the population; a score of 50 is normal in the general population of the United States, with a standard deviation of 10 points (16).

The studies involved in this systematic review were very variable in terms of design, diagnoses, outcome measures and quality assessment. While the process of identifying the final study set was relatively straightforward, classifying studies by region was not. More than 40% of studies involving the upper limb which included the use of CATs required contact with the corresponding author to ascertain the conditions being investigated. It was surprising that many authors could not recall the diagnoses or confirm them after a search of original research documents. The properties of CATs, like all

outcome measures, should be defined for the patients being tested and not for the instruments themselves.[[28]] Any deviation from this principle can compromise the validity of the work and of the measurement system itself. This is particularly relevant, as CATs were originally developed for assessing chronic conditions. The issue of validity is also important when studies are performed on sets of patients with a range of traumatic and non-traumatic conditions. Furthermore, most studies were observational, cross-sectional investigations. Consequently, reliability and measurement error was rarely analyzed. The few longitudinal investigations in this selection often lacked definition of the test conditions and the ‘stability’ of patients around and between the assessment timepoints. The quality of interpretability in these studies was also very variable, curtailing the ability to fully translate our findings.

Early evidence of the positive psychometric properties of CATs in the assessment of patients who have suffered injuries to the upper limbs is compelling. Nevertheless, it is important to recognize the challenges in translating research using CATs into general orthopaedic practice globally. The technology behind CATs is clearly central to their benefits over fixed-length scales. However, aspects concerning resources, regulation, integration and compliance with electronic modes of outcome measurement may provide both actual and perceived barriers to their adoption. A further aspect relates to culture, where many healthcare systems around the world have yet to measure

outcomes routinely in their trauma service, let alone use computers in the process. Language presents a further issue, and cross-cultural translation of CATs is required. Taking this into consideration, all the studies in our review were conducted in a handful of specialist centres in the United States, mostly in follow-up clinics. Ultimately, there is a need for longitudinal studies on groups of trauma patients accounting for the mechanism and severity of injury, the time since injury and clinically important change. More recent studies involving CATs include this level of detail.[[29,30]].

Studies should also be conducted in different geographical settings and at different stages of the care pathway, such as in fracture clinics, and pre and postoperatively during inpatient treatment. Future work should also include quality assessment tools such as the COSMIN checklist during the initial design of the study. This may allow for a standardized approach that improves quality and yields more robust findings.

These observations must be considered in the light of some limitations. Firstly, the broad search strategy generated a high volume of data and consequently a very heterogenous selection of studies. Processing large data sets could lead to errors in retrieving relevant studies. This may affect the internal validity of this work, alongside the issue of publication bias, where unpublished studies recording negative, unfavourable outcomes may exist.[[31]] Secondly, we were also cautious of errors in interpreting findings from a group of studies whose design and characteristics were very heterogeneous. However,

we believe our approach was systematic and justified given that this field of work is relatively new and the studies involved many disciplines including the development of different outcome measures, psychometric evaluation and assessment of health following surgery. Thirdly, although the COSMIN checklist is probably the most appropriate assessment tool for this review, deploying a scoring system on a relatively heterogeneous group of studies may be prone to errors. Error is possible whenever subjective judgements are made to ensure the 'best possible fit'. This may also be compounded by the variability in the level and quality of reporting and potential differences in the terminology between studies and the COSMIN criteria. These were minimized by independent analysis and consensus-based discussions. Finally, there was an insufficient number of similar studies to combine data points and provide a more thorough evaluation of psychometric properties and level of evidence. Given the rise in research using CATs, future work in this field may enable more detailed and quantitative evaluation.

In conclusion, this systematic review aimed to identify and compare the psychometric properties of CATs with fixed-length scales in patients suffering trauma involving the upper limb. Early studies show that CATs are valid and reliable outcome measures, performing as well as if not better than established fixed-length scales in trauma. They also show superior properties, such as improved floor-ceiling effects and user-friendly application. CATs are increasingly being used in outcomes research. Further psychometric evaluation

involving longitudinal studies and groups of patients with specific injuries from different geographical and therapeutic settings are required to assess their properties before adoption as measurements of choice in assessing patient reported outcomes.





# References

1. **Quispe JC, Herbert B, Chadayammuri VP, et al.** Transarticular plating for acute posterior sternoclavicular joint dislocations: a valid treatment option? *Int Orthop* 2016;40:1503–1508.
2. **Pynsent PB, Fairbank JCT, Carr AJ.** *Outcome Measures in Orthopaedics and Orthopaedic Trauma*. Second ed. New York: Oxford University Press, 2004. [[bibmisc]]
3. **Nelson EC, Eftimovska E, Lind C, et al.** Spotlight: Patient Centred Care. Patient reported outcome measures in practice. *BMJ* 2015;350:1–3.
4. **Suk M, Norvell DC, Hanson B, Dettori JR, Helfet D.** Evidence-based orthopaedic surgery: what is evidence without the outcomes? *J Am Acad Orthop Surg* 2008;16:123–129.
5. **Suk M, Hanson BP, Norvell DC, Helfet DL.** *AO Handbook. Musculoskeletal Outcomes Measures and Instruments*. Second ed. Davos: AO Publishing, 2009. [[bibmisc]]
6. **Brodke DJ, Saltzman CL, Brodke DS.** PROMIS for orthopaedic outcomes measurement. *J Am Acad Orthop Surg* 2016;24:744–749.
7. **Beaton DE, Wright JG, Katz JN, Upper Extremity Collaborative Group.** Development of the QuickDASH: comparison of three item-reduction approaches. *J Bone Joint Surg [Am]* 2005;87-A:1038–1046.
8. **Hung M, Hon SD, Franklin JD, et al.** Psychometric properties of the PROMIS Physical Function item bank in patients with spinal disorders. *Spine (Phila Pa 1976)* 2014;39:158–163.
9. **Hung M, Stuart AR, Higgins TF, Saltzman CL, Kubiak EN.** Computerized adaptive testing using the PROMIS Physical Function item bank reduces test burden with less ceiling effects compared with the short musculoskeletal function assessment in orthopaedic trauma patients. *J Orthop Trauma* 2014;28:439–443.
10. **Cella D, Gershon R, Lai JS, Choi S.** The future of outcomes measurement: item

banking, tailored short-forms, and computerized adaptive assessment. *Qual Life Res* 2007;16(Suppl 1):133–141.

11. **Revicki DA, Cella DF.** Health status assessment for the twenty-first century: item response theory, item banking and computer adaptive testing. *Qual Life Res* 1997;6:595–600.
12. **Cella D, Yount S, Rothrock N, et al.** The Patient-Reported Outcomes Measurement Information System (PROMIS): progress of an NIH Roadmap cooperative group during its first two years. *Med Care* 2007;45(5 Suppl 1):S3–S11.
13. **Cook KF, Choi SW, Crane PK, et al.** Letting the CAT out of the bag: comparing computer adaptive tests and an 11-item short form of the Roland-Morris Disability Questionnaire. *Spine (Phila Pa 1976)* 2008;33:1378–1383.
14. **Hays RD, Morales LS, Reise SP.** Item response theory and health outcomes measurement the 21st Century. *Med Care* 2000;38(9 Suppl):II28–II42.
15. **Cella D, Riley W, Stone A, et al.** The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. *J Clin Epidemiol* 2010;63:1179–1194.
16. **Hung M, Clegg DO, Greene T, Saltzman CL.** Evaluation of the PROMIS Physical Function item bank in orthopaedic patients. *J Orthop Res* 2011;29:947–953.
17. **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;61:17–33.
18. **Hung M, Stuart AR, Higgins TF, Saltzman CL, Kubiak EN.** Computerized adaptive testing using the PROMIS Physical Function item bank reduces test burden with less ceiling effects compared with the short musculoskeletal function assessment in orthopaedic trauma patients. *J Orthop Trauma* 2014;28:439–443.
19. **Bruce B, Fries J, Lingala B, Hussain YN, Krishnan E.** Development and assessment of

floor and ceiling items for the PROMIS Physical Function item bank. *Arthritis Res Ther* 2013;15:R144.

20. **Fries J, Rose M, Krishnan E.** The PROMIS of better outcome assessment: responsiveness, floor and ceiling effects, and Internet administration. *J Rheumatol* 2011;38:1759–1764.

21. **Mokkink LB, Terwee CB, Knol DL, et al.** Methodology Protocol of the COSMIN study : COnsensus-based Standards for the selection of health Measurement INstruments. *BMC Med Res Methodol* 2006;6:1–7.

22. **Mokkink LB, Terwee CB, Patrick DL, et al.** The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res* 2010;19:539–549.

23. **Terwee CB, Mokkink LB, Knol DL, et al.** Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res* 2012;21:651–657.

24. **Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group.** Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 2009;6:e1000097.

25. **Wang Y-C, Hart DL, Cook KF, Mioduski JE.** Translating shoulder computerized adaptive testing generated outcome measures into clinical practice. *J Hand Ther* 2010;23:372–382.

26. **Morgan JH, Kallen MA, Okike K, Lee OC, Vrahas MS.** PROMIS Physical Function Computer Adaptive Test compared with other upper extremity outcome measures in the evaluation of proximal humerus fractures in patients older than 60 years. *J Orthop Trauma* 2015;29:257–263.

27. **Okike K, Lee OC, Makanji H, et al.** Comparison of locked plate fixation and

nonoperative management for displaced proximal humerus fractures in elderly patients. *Am J Orthop (Belle Mead NJ)* 2015;44:E106–E112.

28. **Slobogean GP, Slobogean BL.** Measuring shoulder injury function: common scales and checklists. *Injury* 2011;42:248–252.

29. **Beckmann JT, Hung M, Bounsanga J, et al.** Psychometric evaluation of the PROMIS Physical Function Computerized Adaptive Test in comparison to the American Shoulder and Elbow Surgeons score and Simple Shoulder Test in patients with rotator cuff disease. *J Shoulder Elbow Surg* 2015;24:1961–1967.

30. **Nota SPFT, Bot AGJ, Ring D, Kloen P.** Disability and depression after orthopaedic trauma. *Injury* 2015;46:207–212.

31. **Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR.** Publication bias in clinical research. *Lancet* 1991;337:867–872.

32. **Cook KF, Roddey TS, O'Malley KJ, Gartsman GM.** Development of a Flexilevel Scale for use with computer-adaptive testing for assessing shoulder function. *J Shoulder Elbow Surg* 2005;14(1 Suppl S):90S–94S.

33. **Döring AC, Nota SP, Hageman MG, Ring DC.** Measurement of upper extremity disability using the Patient-Reported Outcomes Measurement Information System. *J Hand Surg Am* 2014;39:1160–1165.

34. **Haley SM, Ni P, Jette AM, et al.** Replenishing a computerized adaptive test of patient-reported daily activity functioning. *Qual Life Res* 2009;18:461–471.

35. **Hart DL, Cook KF, Mioduski JE, Teal CR, Crane PK.** Simulated computerized adaptive test for patients with shoulder impairments was efficient and produced valid measures of function. *J Clin Epidemiol* 2006;59:290–298.

36. **Hart DL, Wang Y-C, Cook KF, Mioduski JE.** A computerized adaptive test for patients with shoulder impairments produced responsive measures of function. *Phys Ther*

2010;90:928–938.

37. **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;94:2291–2296.

38. **Janssen SJ, Ter Meulen DP, Nota SP, Hageman MG, Ring D.** Does verbal and nonverbal communication of pain correlate with disability? *Psychosomatics* 2015;56:338–344.

39. **Jayakumar P, Overbeek CL, Ring DC.** Relationship of age on enjoyment of physical activity in upper extremity illness. *Hand* 2015;10:767–772.

40. **Jette AM, Haley SM, Tao W, et al.** Prospective evaluation of the AM-PAC-CAT in outpatient rehabilitation settings. *Phys Ther* 2007;87:385–398.

41. **Kortlever JTP, Janssen SJ, van Berckel MM, Ring D, Vranceanu AM.** What is the most useful questionnaire for measurement of coping strategies in response to nociception? *Clin Orthop Relat Res* 2015;473:3511–3518.

42. **Lehman LA, Woodbury M, Shechtman O, et al.** Development of an item bank for a computerised adaptive test of upper-extremity function. *Disabil Rehabil* 2011;33:2092–2104.

43. **Mellema JJ, O'Connor CM, Overbeek CL, Hageman MG, Ring D.** The effect of feedback regarding coping strategies and illness behavior on hand surgery patient satisfaction and communication: a randomized controlled trial. *Hand* 2015;10:503–511.

44. **Menendez ME, Bot AG, Hageman MG, et al.** Computerized adaptive testing of psychological factors: relation to upper-extremity disability. *J Bone Joint Surg [Am]* 2013;95-A:e149.

45. **Menendez ME, Mudgal CS, Jupiter JB, Ring D.** Health literacy in hand surgery patients: a cross-sectional survey. *J Hand Surg Am* 2015;40:798–804.e2.

46. **Menendez ME, Eberlin KR, Mudgal CS, Ring D.** Language barriers in Hispanic

- patients: relation to upper-extremity disability. *Hand* 2015;10:279–284.
47. **Menendez ME, Chen NC, Mudgal CS, Jupiter JB, Ring D.** Physician empathy as a driver of hand surgery patient satisfaction. *J Hand Surg Am* 2015;40:1860–1865.e2.
  48. **Moradi A, Mellema JJ, Oflazoglu K, et al.** The relationship between catastrophic thinking and hand diagram areas. *J Hand Surg Am* 2015;40:2440–2446.e5.
  49. **Overbeek CL, Nota SPFT, Jayakumar P, Hageman MG, Ring D.** The PROMIS physical function correlates with the QuickDASH in patients with upper extremity illness. *Clin Orthop Relat Res* 2014;473:311–317.
  50. **Peters RM, Menendez ME, Mellema JJ, Ring D, Vranceanu AM.** Sleep disturbance and upper-extremity disability. *Arch Bone Jt Surg* 2016;4:35–40.
  51. **Sindhu BS, Lehman LA, Tarima S, et al.** Influence of fear-avoidance beliefs on functional status outcomes for people with musculoskeletal conditions of the shoulder. *Phys Ther* 2012;92:992–1005.
  52. **Sindhu BS, Wang Y-C, Lehman LA, Hart DL.** Differential item functioning in a Computerized Adaptive Test of Functional Status for people with shoulder impairments is negligible across pain intensity, gender, and age groups. *Occup Particip Heal* 2013;33:86–99.
  53. **Stuart AR, Higgins TF, Hung M, et al.** Reliability in measuring preinjury physical function in orthopaedic trauma. *J Orthop Trauma* 2015;29:527-532.
  54. **Tyser AR, Beckmann J, Franklin JD, et al.** Evaluation of the PROMIS Physical Function Computer Adaptive Test in the upper extremity. *J Hand Surg Am* 2014;39:2047–2051.e4.
  55. **van Leeuwen W, van der Vliet Q, Janssen S, et al.** Does perceived injustice correlate with pain intensity and disability in orthopaedic trauma patients? *Injury* 2016;47:1212-1216.