

■ TRAUMA

Are validated outcome measures used in distal radial fractures truly valid?

A CRITICAL ASSESSMENT USING THE CONSENSUS-BASED STANDARDS FOR THE SELECTION OF HEALTH MEASUREMENT INSTRUMENTS (COSMIN) CHECKLIST



Y. V. Kleinlugtenbelt,
R. W. Nienhuis,
M. Bhandari,
J. C. Goslings,
R. W. Poolman,
V. A. B. Scholtes

*Joint Research OLVG
East, Amsterdam, The
Netherlands*

■ Y. V. Kleinlugtenbelt, MD, Orthopaedic Surgeon, Department of Orthopaedic and Trauma Surgery, Deventer Ziekenhuis, Nico Bolkesteinlaan 75 7416 SE Deventer, The Netherlands.
■ R. W. Nienhuis, MD, Orthopaedic Resident, Department of Orthopaedic and Trauma Surgery, Antonius Ziekenhuis, Koekoekslaan 1, 3435 CM, Nieuwegein, The Netherlands.
■ M. Bhandari, FRCSC, MD, PhD, Orthopaedic Surgeon, Professor, Division of Orthopaedic Surgery, McMaster University, Hamilton, Ontario, Canada.
■ J. C. Goslings, MD, PhD, Trauma Surgeon, Professor, Trauma Unit, Department of Surgery, Academic Medical Center, University of Amsterdam, P.O. Box 22660, 1100, DD, Amsterdam, The Netherlands.
■ R. W. Poolman, MD, PhD, Orthopaedic Surgeon, Department of Orthopaedic and Trauma Surgery.
■ V. A. B. Scholtes, PhD, Program Manager, Department of Orthopaedic and Trauma Surgery, JointResearch OLVG East, Oosterpark 9, P.O. Box 95500, 1090 HM, Amsterdam, The Netherlands.

Correspondence should be sent to Dr Y. V. Kleinlugtenbelt; e-mail: ijdoklb@hotmail.com

doi:10.1302/2046-3758.54.2000462

Bone Joint Res 2016;5:153–161.

Received: 20 July 2015;

Accepted: 02 February 2016

Objectives

Patient-reported outcome measures (PROMs) are often used to evaluate the outcome of treatment in patients with distal radial fractures. Which PROM to select is often based on assessment of measurement properties, such as validity and reliability. Measurement properties are assessed in clinimetric studies, and results are often reviewed without considering the methodological quality of these studies. Our aim was to systematically review the methodological quality of clinimetric studies that evaluated measurement properties of PROMs used in patients with distal radial fractures, and to make recommendations for the selection of PROMs based on the level of evidence of each individual measurement property.

Methods

A systematic literature search was performed in PubMed, Embase, CINAHL and PsycINFO databases to identify relevant clinimetric studies. Two reviewers independently assessed the methodological quality of the studies on measurement properties, using the Consensus-based Standards for the selection of health Measurement Instruments (COSMIN) checklist. Level of evidence (strong / moderate / limited / lacking) for each measurement property per PROM was determined by combining the methodological quality and the results of the different clinimetric studies.

Results

In all, 19 out of 1508 identified unique studies were included, in which 12 PROMs were rated. The Patient-rated wrist evaluation (PRWE) and the Disabilities of Arm, Shoulder and Hand questionnaire (DASH) were evaluated on most measurement properties. The evidence for the PRWE is moderate that its reliability, validity (content and hypothesis testing), and responsiveness are good. The evidence is limited that its internal consistency and cross-cultural validity are good, and its measurement error is acceptable. There is no evidence for its structural and criterion validity. The evidence for the DASH is moderate that its responsiveness is good. The evidence is limited that its reliability and the validity on hypothesis testing are good. There is no evidence for the other measurement properties.

Conclusion

According to this systematic review, there is, at best, moderate evidence that the responsiveness of the PRWE and DASH are good, as are the reliability and validity of the PRWE. We recommend these PROMs in clinical studies in patients with distal radial fractures; however, more clinimetric studies of higher methodological quality are needed to adequately determine the other measurement properties.

Cite this article: *Bone Joint Res* 2016;5:153–161

Keywords: PROM; Distal radius fracture; COSMIN; Validation study

Article focus

- The aim of this systematic review was to evaluate the methodological quality of the clinimetric studies that evaluated measurement properties of the available patient reported outcome measures (PROMs) used in patients with distal radial fractures.
- To determine which PROM, based on the level of evidence of each individual measurement property, is most appropriate for the evaluation of patients with distal radial fractures.

Key findings

- The two PROMs that were most extensively evaluated were the patient rated wrist evaluation (PRWE) (with seven of nine measurement properties investigated) and the Disabilities of Arm, Shoulder and Hand (DASH) (with four of nine investigated). The methodological quality of these studies ranged at best from poor to good.

Key messages

- Strong evidence supporting ‘good quality’ of any of the current available PROMs in patients with distal radial fractures is lacking.
- The PRWE and DASH are the two most extensively evaluated PROMs. Their measurement properties were mainly good but the methodological quality of the clinimetric studies was low; this does mean that these results may be biased.
- For now we recommend to use the PRWE or DASH, but more clinimetric studies of higher methodological quality are needed to select PROMs in patients with distal radius fractures with greater confidence.

Strengths and limitations

- Strength: This is the first study that has used the COnsensus-based Standards for the Selection of Health Measurement INstruments (COSMIN) checklist to systematically review the methodological quality of studies on the measurement properties of PROMs in the evaluation of treatment of distal radial fractures.
- Strength: Our search was not just limited to English language studies, as both reviewers have a good knowledge of German and Dutch.
- Limitation: It was not possible to distinguish between poor study reporting and poor methodological quality.

Introduction

Distal radial fractures account for approximately 17 % of all fractures¹ and the distal radius is the most common fracture site in the upper extremity.²⁻⁴ Despite its high incidence, there is no treatment consensus for these fractures.⁵ To conduct best evidence clinical trials in distal radial fracture treatment, and to properly compare trial results, there must be consensus on the use of outcome measures. Historically,

outcome assessment after distal radial fractures focused on imaging and physical examination (e.g. grip strength and range of motion). These assessments, however, do not represent the patients’ perspective as they do not take the patients’ feelings, opinion or wellbeing into account, which are likely to be more important for the patient.⁶

In the last two decades, outcomes assessment has shifted towards a patient-centred approach. This approach assesses the outcome based directly on the opinion of the patient. Outcomes such as pain and functional ability, which are highly relevant for patients, can be assessed by patient-reported outcome measures (PROMs).⁷

Currently, a wide variety of PROMs are available and are used to assess patient-reported functional outcomes for upper limb and wrist disorders.⁸⁻²⁰ Several (non-)systematic studies have reviewed the existing literature in order to present available PROMs for assessing wrist and hand function in general.²¹⁻²⁵ Over a period of 25 years, the two most extensively used PROMs for evaluating the treatment outcome of patients with distal radial fractures.²⁶ were the Disabilities of Arm, Shoulder and Hand (DASH), and the (original or modified) Gartland and Werley scoring system. However, the patient-rated wrist evaluation (PRWE) was found to have the best measurement properties, e.g. it was found to be the most reliable, valid and responsive instrument for these patients. This conclusion was based on the results of the available clinimetric studies.²⁶ Clinimetrics is a scientific discipline that aims to develop methods of assessing the properties of health measurement instruments, with the aim of improving the quality of outcome measures. Although the measurement properties were found to be good, the authors did not incorporate the methodological quality of these clinimetric studies.

It is important for the understanding of this systematic review to distinguish between the ‘methodological quality’ of clinimetric studies on PROMs and the ‘quality’ (e.g. the measurement properties) of the PROMs themselves. Evidently a PROM is only as good as the methodological quality of its study. In order to assess the methodological quality of clinimetric studies (i.e. studies on measurement properties) on PROMs, the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) group formulated a set of guidelines. First, the COSMIN group reached consensus on terminology, definitions and a taxonomy of measurement properties of PROMs in an international Delphi study. Next, the group developed a checklist containing standards for evaluating the methodological quality of studies on the measurement properties (e.g., reliability) of measurement instruments (e.g. DASH) (www.cosmin.nl).²⁷ The best PROM should have a high level of evidence (e.g. as evaluated in high quality studies) supporting good quality on all measurement properties. The definitions and a description of the measurement properties are given in Table I.

Table I. Definitions of the measurement properties.

Definitions of the measurement properties	
Internal consistency	The degree of the interrelatedness among the items “Do the different questions in a PROM that are meant to measure the same general construct produce similar scores?”
Reliability	The proportion of the total variance in the measurements which is because of “true” differences among patients “How close are repeated measurements?”
Measurement error	The systematic error and random error of a patient’s score that are not attributed to true changes in the construct to be measured “What amount of change in a score cannot be considered a real or true change?”
Content validity	The degree to which the content of a health-related patient-reported outcomes (HR-PRO) instrument is an adequate reflection of the construct to be measured “Are all items relevant for the specific population and have important activities been missed?”
Structural validity	The degree to which the scores of an HR-PRO instrument are an adequate reflection of the dimensionality of the construct to be measured “Do all items in a PROM reflect single or multiple constructs?”
Hypotheses testing	The degree to which the scores of an HR-PRO instrument are consistent with hypotheses (for instance with regard to internal relationships, relationships to scores of other instruments, or differences between relevant groups) based on the assumption that the HR-PRO instrument validly measures the construct to be measured “What is the expected relationship with other PROMs assessing comparable constructs?”
Cross-cultural validity	The degree to which the performance of the items on a translated or culturally adapted HR-PRO instrument is an adequate reflection of the performance of the items of the original version of the HR-PRO instrument “Has the PROM been correctly translated and retested in another language and cultural setting?”
Criterion validity	The degree to which the scores of an HR-PRO instrument are an adequate reflection of a “gold standard” “Is the PROM tested against the benchmark PROM?”
Responsiveness	The ability of an HR-PRO instrument to detect change over time in the construct to be measured “If patients improve or worsen over time does this change in the PROM accordingly?”
Interpretability*	The degree to which one can assign qualitative meaning—that is, clinical or commonly understood connotations—to an instrument’s quantitative scores or change in scores “What do the scores or change in scores of a PROM mean?”

Clarification in bold

*Is not a real measurement property, but nevertheless it is a meaningful requirement for the applicability of PROMs in research
PROM, patient-reported outcome measure

The aim of this systematic review was to evaluate the methodological quality (using the COSMIN checklist) of the clinimetric studies that evaluated measurement properties of the available PROMs used in patients with distal radial fractures, and to make recommendations for the selection of PROMs based on the level of evidence of each individual measurement property. The results of this study might help us to determine which PROM is most appropriate for the evaluation of patients with distal radial fractures.

Materials and Methods

Literature search. We performed a literature search on November 13, 2015 to identify all published studies on the measurement properties of PROMs in the evaluation of treatment of distal radial fractures. The following databases were searched with specific index terms and derivatives of these terms: PubMed (1990 to 2015), Embase (1990 to 2015), CINAHL (1990 to 2015), and PsycINFO (1990 to 2015). In PubMed we used a validated search filter for finding studies on measurement properties.²⁸ We also added the names of all PROMs that are described for wrist disorders.²⁹ The full search strategy is provided in the supplementary material. We restricted our search to studies published in English, German and Dutch because both reviewers are fluent in these languages. Reference lists were hand-searched to identify additional relevant studies.

Selection criteria. Two reviewers (YK and RN) independently assessed all titles and abstracts. We included studies with a description of the measurement properties of PROMs used in patients with a distal radial fracture. When in doubt about the applicability of a study, the full text article was retrieved and screened for eligibility. Afterwards, the researchers discussed their assessments and consensus was reached. In cases where consensus couldn’t be obtained, a third reviewer (VS), was employed to achieve consensus.

Assessment of the quality of the studies. The same two reviewers independently rated the methodological quality of the studies using the COSMIN checklist (www.cosmin.nl).³⁰

The COSMIN checklist consists of 11 separate checklists, called “boxes”. In nine boxes the quality of nine measurement properties is addressed: a) internal consistency, b) reliability, c) measurement error, d) content validity, e) structural validity, f) hypotheses testing, g) cross-cultural validity, h) criterion validity and i) responsiveness. The last box, “j: interpretability”, is not a measurement property, but nevertheless it is a meaningful requirement for the applicability of PROMs in research. The generalisability of the results is determined with a final box. The definitions of the measurement properties and interpretability are given in Table I.

In each box, the methodological quality can be evaluated based on a variety of items addressing adequate study

design and statistical analysis. Each question in any box must be rated as 'excellent', 'good', 'fair', 'poor' or 'not applicable'. Scoring is then performed using the criteria set by the COSMIN group. To obtain a total score for the methodological quality of one of the boxes, "the worst score counts" algorithm was applied as set out by the COSMIN guidelines,³¹ meaning that the methodological quality of that measurement property was only rated 'excellent' if all relevant questions pertaining to that box (e.g. measurement property) were scored as excellent. In all boxes, a small sample size was considered poor methodological quality. As a rule of thumb, a sample size of ≥ 100 received a rating of 'excellent', 50 to 100 received 'good', 30 to 50 was rated 'fair', and less than 30 was rated as 'poor'.³¹

Level of evidence of the measurement properties per PROM. For each PROM, we determined the level of evidence by combining the results of the different studies for each measurement property, as described by Terwee et al.³¹ The following factors were taken into account: the number of studies (one or multiple), the methodological quality of the studies (excellent/good/fair/poor/not available), and consistency of the results (positive/negative). Based on these factors each measurement property per PROM could be ranked as strong, moderate, limited or conflicting evidence. Only when the methodological quality of the clinimetric study/studies was poor was the level of evidence rated as 'unknown'.

Results

Included studies. A total of 2064 studies were retrieved by the electronic search performed in PubMed ($n = 720$), EMBASE ($n = 1075$) and CINAHL/ PsycINFO ($n = 269$) (Fig. 1). After removing duplicates, 1508 unique studies were identified. The titles and abstracts were independently screened by two researchers, after which 27 studies were deemed potentially eligible. After retrieving and reading the full text, 19 studies were included. Reference evaluation of these 19 articles did not yield any additional relevant studies.

Overall results. In the 19 included studies, a total of 12 PROMs were evaluated (Table II). In three papers, multiple PROMs were evaluated: three,³² three³³ and five,³⁴ respectively. Most studies (80%) evaluated more than one measurement property. None of the studies evaluated structural validity. Criterion validity was also not evaluated in any of the studies. However, this was expected given that there are no measurement instruments that can be used as a benchmark, which is a prerequisite of this measurement property. A complete overview of the study characteristics is shown in Table III.

Of all PROMs, the PRWE has been studied most extensively, followed by the DASH. The eight studies evaluating the PRWE assessed almost all measurement properties: seven of the nine (Table IV). However, overall, the methodological quality of these studies was low, varying from poor to fair for internal consistency,

reliability, measurement error, cross-cultural validity and responsiveness, and varying from poor to good for content validity and hypothesis testing. Interpretability was also assessed, but these studies were of poor methodological quality.

The four studies evaluating the DASH^{32,34-36} assessed less than half of the measurement properties: four of nine. The methodological quality of these studies was generally low, varying from persistently poor for internal consistency, poor to fair for reliability, and consistently fair for responsiveness. Measurement error, content validity, hypothesis testing, cross-cultural validity and interpretability were not assessed in any of the studies (Table IV).

Of the other ten PROMs, one to three measurement properties were assessed. These concerned mostly internal consistency, reliability and responsiveness. Overall, the methodological quality of these clinimetric studies was at best poor to fair (Table V). This is mainly due to the low sample size in the majority of these studies but can also be secondarily attributed to the high amount of items that were scored as "not applicable". Finally, the lack of description surrounding the statistical methods that were used also contributed to the poor rating.

Level of evidence of the measurement properties per PROM. The synthesis of results per PROM and their accompanying level of evidence are presented in Table VI.

The highest levels of evidence were found for the measurement properties of the PRWE. Nevertheless, the evidence is, at best, limited to moderate. For instance, reliability (assessed in 78% of the studies) ranged from 0.81 to 0.97 (Table VI). Three studies were of poor methodological quality, and four were of fair quality (Table IV). Therefore, the synthesis of these results is that there is moderate evidence supporting good reliability. There is also moderate evidence that the validity (content and hypothesis testing) and responsiveness are good. The evidence is limited in that its internal consistency and cross-cultural validity are good, and its measurement error is acceptable. There is no evidence for its structural and criterion validity. The evidence for the DASH is moderate that its responsiveness is good. The evidence is limited that its reliability and the validity on hypotheses testing are good. There is no evidence for the other measurement properties. The evidence for the other ten PROMs is mainly unknown, since the quality of the studies that evaluated some of the PROM measurement properties (mainly internal consistency, reliability and/or responsiveness) was mainly poor methodologically.

Discussion

The aim of this systematic review was to evaluate the methodological quality of the clinimetric studies that evaluated measurement properties of the available PROMs used in patients with distal radial fractures, and to make recommendations for the selection of PROMs based on the level of evidence of each individual measurement property.

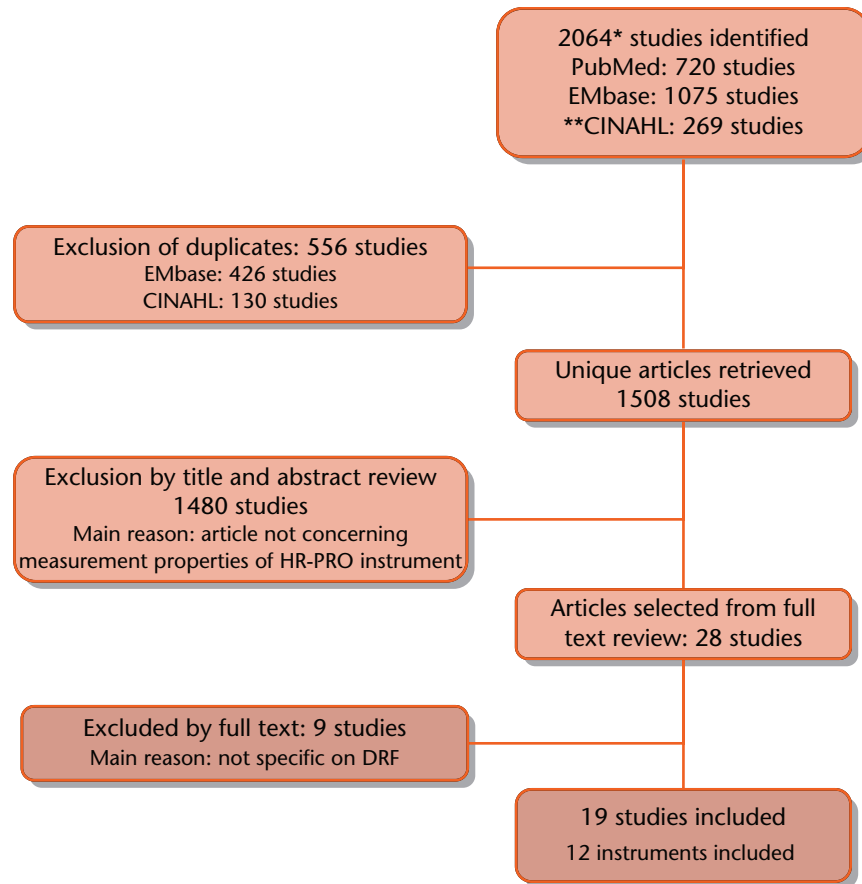


Fig. 1

Search strategy and selection of articles. *Nov 13, 2015. **Cinahl search includes PsycInfo database. HR-PRO, health-related patient-reported outcomes; DRF, distal radial fracture.

Table II. Patient-related outcome instruments included in the review^{8-14,16-20}

Abbreviation	Full name	Original author
PRWE	Patient-Rated Wrist Evaluation	MacDermid ⁹
DASH	Disabilities of Arm, Shoulder and Hand	Hudak ⁸
MHQ	Michigan Hand Questionnaire	Chung ¹¹
SF-36	Short Form-36	Ware ¹²
PEM	Patient Evaluation Measure	Macey ¹⁰
AIMS2	Arthritis Impact Measurement Scale	Meenan ¹⁴
BWH-CTQ	Brigham and Women’s Hospital Carpal Tunnel Questionnaire	Levine ¹³
IOF-WFQ	International Osteoporosis Foundation Wrist Fracture Questionnaire	Lips ¹⁶
PFW	Patient Focused Wrist Outcome Instrument	Bialocerkowski ¹⁷
TSK	Tampa Scale of Kinesophobia	Kori ¹⁸
CAT	Catastrophizing Subscale of the Coping Strategies Questionnaire	Rosenstiel ¹⁹
SES	Self-Efficacy Scale	Altmaier ²⁰

Key findings. The two PROMs that were most extensively evaluated were the PRWE (with seven of nine measurement properties investigated) and the DASH (with four of nine investigated). The methodological quality of these studies ranged at best from poor to good. Therefore, after synthesis of the scores and incorporating the levels

of evidence, the quality of these two PROMS is not supported with strong levels of evidence on any of the measurement properties. For the PRWE, there is moderate evidence supporting good reliability, content validity, hypotheses testing and responsiveness. The evidence is only limited in that the measurement error is acceptable and the cross-cultural validity and internal consistency are good. Structural validity and criterion validity were never evaluated, so these lack in evidence. The evidence for interpretability, which is not a measurement property, is unknown, since this was only evaluated in three studies with poor methodological quality. The DASH showed at best moderate evidence for good responsiveness and limited evidence for good hypotheses testing and reliability. All other measurement properties were found to be lacking in evidence.

These findings do not mean that these and other PROMs have poor measurement properties and thus are of poor quality. Since we found that, overall, the measurement properties were good but the methodological quality of the clinimetric studies was low, it does mean that these results may be biased. Therefore, the results of our review do imply that studies of higher methodological quality are needed to properly assess their measurement

Table III. Study characteristics³²⁻⁴⁹

Measurement Instrument	Study	n	Mean age	Gender	Country	Language
			(range or sd)	Male (%)		
Patient-Rated Wrist Evaluation	Gabl ³⁷	133	62 (19 to 92)	27	Austria	German*
	Hemelaers ³⁸	44	56 (15)	36	Switzerland	German
	MacDermid ³⁹	36/101	45 (10) / 50 (16)	33 / 31	Canada	English*
	MacDermid ³²	59	53 (18)	37	Canada	English*
	Wilcke ³⁵	99	58 (18)	20	Sweden	Swedish
	Lovgren ³⁴	16	52 (12)	19	Sweden	Swedish
	Mehta ⁴⁰	50	46 (14)	56	India	Hindi
	Kim ⁴¹	63	56 (19 to 83)	27	Rep. Korea	Korean
	Schonnemann ⁴²	60/29	55 (19 to 86)	27	Denmark	Danish
	Walenkamp ⁴³	102	59 (48 to 66)	30	Netherlands	Dutch
Disabilities of Arm, Shoulder and Hand	Macdermid ³²	59	53 (18)	37	Netherlands	English*
	Westphal ³⁶	107	59 (17 to 84)	27	Germany	German
	Westphal ⁴⁴	72	60 (16)	29	Germany	German
	Lovgren ³⁴	16	52 (12)	19	Sweden	Swedish
Michigan Hand Questionnaire	Kotsis ⁴⁵	47 / 37	48 (17) / 51(16)	32 / 38	USA	English
	Shauver ⁴⁶	51	50 (19 to 83)	37	USA	English
	Waljee ⁴⁷	128	61 (9)	27	USA/UK	English*
	Amadio ³³	21	57 (14 to 84)	14	USA	English*
Short Form-36	MacDermid ³²	59	53 (18)	37	Canada	English*
	Forward ⁴⁸	200	54 (24 to 80)	36	UK	English*
	Amadio ³³	21	57 (14 to 84)	14	USA	English*
Patient Evaluation Measure Arthritis Impact Measurement Scale2	Amadio ³³	21	57 (14 to 84)	14	USA	English*
	Amadio ³³	21	57 (14 to 84)	14	USA	English*
Brigham and Women's Hospital Carpal Tunnel Questionnaire	Amadio ³³	21	57 (14 to 84)	14	USA	English*
International Osteoporosis Foundation Wrist Fracture Questionnaire	Lips ¹⁶	105	63 (8)	12	UK/NL/Ita/BE	English/Dutch/Italian*
Patient Focused Wrist Outcome Instrument	Bialocerkowski ⁴⁹	26	62 (22 to 84)	15	Australia	English
Tampa Scale of Kinesophobia	Lovgren ³⁴	16	52 (12)	19	Sweden	Swedish
Catastrophizing Subscale of the Coping Strategies Questionnaire	Lovgren ³⁴	16	52 (12)	19	Sweden	Swedish
Self-Efficacy Scale	Lovgren ³⁴	16	52 (12)	19	Sweden	Swedish

*It can be deduced as per the COnsensus-based Standards for the selection of health Measurement INstruments guidelines, often the country in which the study is performed and the language version of the measurement instrument that was used are not mentioned explicitly, but can be deduced from the affiliation of the authors

Table IV. Summary of methodological quality of the studies on measurement properties of the PRWE and DASH³²⁻⁴⁴

	PRWE ³⁷	PRWE ³⁸	PRWE ³⁹	PRWE ³²	PRWE ³⁵	PRWE ³⁴	PRWE ⁴⁰	PRWE ⁴²	PRWE ⁴¹	PRWE ⁴³	DASH ³²	DASH ³⁶	DASH ⁴⁴	DASH ³⁴
Generalisability	Fair	Fair	Fair	Poor	Fair	Excel	Poor	Fair	Fair	Fair	Poor	Fair	Good	Excel
Internal Consistency	Poor	Poor			Fair	Poor	Poor	Poor	Poor	Poor		Poor	Poor	Poor
Reliability		Fair	Poor		Fair	Poor	Fair	Poor	Fair	Poor			Fair	Poor
Measurement Error									Fair	Poor				
Content validity			Fair					Good						
Structural validity														
Hypotheses testing		Fair			Good		Fair	Fair	Poor			Fair		
Cross-cultural					Fair		Poor	Poor	Poor					
Criterion validity														
Responsiveness			Fair	Fair	Fair		Fair	Fair	Poor		Fair		Fair	
Interpretability					Poor				Poor		Poor	Poor		

A full overview of all the scores are shown in the supplementary material

properties. For instance, many PROMs are translated into multiple languages. The PRWE has been correctly translated into 14 languages, following the translation process described by Beaton et al.⁵⁰ Nevertheless, we only found cross-cultural validity studies for the Swedish, Hindi, Korean and Danish versions, because the other translated versions were not adequately evaluated. However, our search was limited to English, German and Dutch, so it can be assumed that the cross-cultural validity was evaluated but the results were not published in any of these languages.

Comparison of results with previous literature. Previous reviews described a variety of PROMS measuring wrist and/or hand disorders in general, but not PROMs specific to distal radial fractures. Goldhahn et al²⁵ advise using a combination of a disease-specific PROM (PRWE), an extremity-specific PROM (DASH) and a generic PROM (SF-36). Changulani et al²² compared the measurement properties of four PROMs for wrist and hand disorders. They concluded that the PRWE is the most responsive instrument for evaluating outcomes in patients with a distal radial fracture. These conclusions were drawn

Table V. Summary of methodological quality of the studies on measurement properties of the other measurement instruments. A full overview of all the scores are shown in the supplementary material^{16,32-34,45,46,48,49}

	MHQ ⁴⁵	MHQ ⁴⁶	MHQ ⁴⁶	SF-36 ³²	SF-36 ³³	PEM ⁴⁸	IOF ¹⁶	PFW ⁴⁹	AIMS2 ³³	BWH ³³	TSK ³⁴	CAT ³⁴	SES ³⁴
Generalisability	Fair	Fair	Poor	Poor	Fair	Poor	Fair	Fair	Fair	Fair	Excellent	Excellent	Excellent
Internal Consistency						Poor	Poor				Poor	Poor	Poor
Reliability							Poor				Poor	Poor	Poor
Measurement Error													
Content validity													
Structural validity													
Hypotheses testing						Poor		Poor					
Cross-cultural													
Criterion validity													
Responsiveness	Fair	Fair	Fair	Fair	Poor		Fair	Poor	Poor	Poor			
Interpretability		Fair											

MHQ, Michigan Hand Questionnaire; SF-36, Short Form-36; PEM, Patient Evaluation Measure; AIMS2, Arthritis Impact Measurement Scale; BWH-CTQ, Brigham and Women’s Hospital Carpal Tunnel Questionnaire; IOF-WFQ International Osteoporosis Foundation Wrist Fracture Questionnaire; PFW, Patient Focused Wrist Outcome Instrument; TSK, Tampa Scale of Kinesophobia; CAT, Catastrophizing Subscale of the Coping Strategies Questionnaire; SES Self-Efficacy Scale

Table VI. Ratings of measurement properties and interpretability of measurement instruments with level of evidence³²⁻⁴⁹

	PRWE ^{32,34,35,37-43}	DASH ^{32,34,36,44}	MHQ ⁴⁵⁻⁴⁷	SF-36 ^{32,33}	PEM ⁴⁸	AIMS2 ³³	BWH ³³	IOF ³³	PFW ⁴⁹	TSK ³⁴	CAT ³⁴	SES ³⁴
Reliability												
Internal consistency	+	?			?		?		?	?	?	?
Cronbach’s alpha	0.89 to 0.97	0.93 to 0.98			0.94		0.96		0.68 to 0.88 to 0.79 to 0.82 0.97 0.95			
Reliability	++	+					?		?	?	?	?
Intraclass correlation coefficient	0.81 to 0.97	0.78 to 0.95					NA		0.81 to 0.85 to 0.57 to 0.84 0.89 0.86			
Measurement error	+											
Smallest detectable change	4.4 to 11.0											
Validity												
Content validity	++											
Structural validity												
Hypotheses testing	++	+			?			+				
Comparator instrument	DASH	Gartland			NA			NA				
Cross-cultural	+											
Criterion validity												
Responsiveness												
Responsiveness	++	++	++	+		?	?	+	?			
Standardised response mean	NA	NA	NA	NA		NA	NA	NA	NA			
INTERPRETABILITY												
Interpretability	?		-									
Minimal important change	11.5											

+ ++ or --- multiple studies of good quality OR 1 study of excellent quality: strong evidence positive/negative result

+ + or -- multiple studies of fair quality OR 1 study of good quality: moderate evidence positive/negative result

+ or - 1 study of fair quality: limited evidence positive/negative result

+ / - conflicting findings

? only studies of poor quality: unknown, due to poor methodological quality

NA, not available (not performed or described)

PRWE, Patient-Rated Wrist Evaluation; DASH, Disabilities of Arm, Shoulder and Hand; MHQ, Michigan Hand Questionnaire; SF-36, Short Form-36; PEM, Patient Evaluation Measure; AIMS2, Arthritis Impact Measurement Scale; BWH-CTQ, Brigham and Women’s Hospital Carpal Tunnel Questionnaire; IOF-WFQ International Osteoporosis Foundation Wrist Fracture Questionnaire; PFW, Patient Focused Wrist Outcome Instrument; TSK, Tampa Scale of Kinesophobia; CAT, Catastrophizing Subscale of the Coping Strategies Questionnaire; SES Self-Efficacy Scale

before the COSMIN checklist was available. The methodological quality of the clinimetric studies was not taken into account and therefore these results may be biased, especially since in the current review we found that the methodological quality of these studies was, at best, fair. Therefore, we can only conclude that the good responsiveness of the DASH and PRWE is supported by moderate evidence.

Hoang-Kim et al²¹ assessed the quality of reviews published on currently used PROMs for assessing function of

the hand and wrist joints. Although they used COSMIN’s taxonomy, terminology and definitions to define the different measurement properties, they did not systematically review the methodological quality of these studies. Nevertheless, they concluded that the PRWE has good construct validity and responsiveness, and found this to be only slightly better than the DASH for assessing patients with wrist injuries. Based on the results of our review we agree that the PRWE is slightly better investigated than the DASH, but disagree with their rating of “good” on some

measurement properties. This difference may be due to the fact that we incorporated the methodological quality of these studies by using the COSMIN checklist instead of only using the COSMIN taxonomy.

Study strengths. To our knowledge, this is the first study that has used the COSMIN checklist to systematically review the methodological quality of studies on the measurement properties of PROMs in the evaluation of treatment of distal radial fractures. Furthermore, the quality of each study was assessed by two independent reviewers, as recommended by the COSMIN group, and a third reviewer in cases of disagreement. Using these methods, we were able to minimise subjective judgement on the outcome. We searched for relevant articles from 1990 onwards, so we consider it unlikely that any relevant PROMs were missed. This is especially true since most PROMs were developed after 1990. Since we found 19 studies eligible from a possible 1508, this shows that our search strategy was very broad and inclusive. Yet, it also demonstrates that the literature on this topic is somewhat lacking. Our search was not just limited to the English language, as both reviewers have a good knowledge of German and Dutch.

Study weaknesses. There were some limitations to this review. As in all reviews, publication bias from unpublished studies may threaten the internal validity as unpublished studies are more likely to report negative or unfavourable results.⁵¹ Another limitation of this study was that it was not always clear to the reviewers if specific methodological aspects were not reported or not performed, making it impossible to distinguish between poor study reporting and poor methodological quality. We did not contact the authors of the studies to clarify these issues. It can be assumed that some studies have been executed properly but are not sufficiently well described according to the COSMIN criteria. This may have affected the quality ratings.

The shortcomings of outcome measurement research in distal radial fractures exposed by this review should not be generalised to all clinimetric research in orthopaedic surgery. However, it is known that strong evidence supporting good quality of multiple PROMs for various pathology is lacking⁵²⁻⁵⁴ so we advise the reader to be cautious when choosing a PROM based on the results of clinimetric studies without considering their methodological quality.

For future research, we believe that it is especially important to further evaluate the measurement properties and interpretability of the PRWE and DASH outcome measures in higher quality studies. Based on the results of the available clinimetric studies, there is no evidence that these PROMs are not useful in evaluating the treatment of distal radial fractures, and therefore we do not believe that it is necessary to develop new instruments. Currently, based on best available evidence, we recommend using the PRWE or DASH to evaluate the outcome of treatment in patients with distal radial fractures but we cannot

stress strongly enough that more clinimetric studies of higher methodological quality are needed in order to more confidently select appropriate PROMs.

According to this systematic review, strong evidence supporting 'good quality' of any of the current available PROMs in patients with distal radial fractures is lacking. The evidence that the responsiveness of the PRWE and DASH is good is moderate, as is the evidence for good validity and reliability of the PRWE. We therefore recommend these PROMs in clinical studies in patients with distal radial fractures; however, more clinimetric studies of higher methodological quality are needed to adequately determine their other measurement properties. If the methodological quality of clinimetric studies continues to increase, PROMs can be selected with greater confidence.

Supplementary material



The full search strategy is provided in supplementary material 1. A full overview of the scores of methodological quality of the studies on measurement properties of all PROMs are shown in Supplementary Tables i to iv.

References

1. Court-Brown CM, Caesar B. Epidemiology of adult fractures: A review. *Injury* 2006;37:691-697.
2. Alffram PA, Bauer GC. Epidemiology of fractures of the forearm. A biomechanical investigation of bone strength. *J Bone Joint Surg [Am]* 1962;44-A:105-114.
3. Dóczy J, Renner A. Epidemiology of distal radius fractures in Budapest. A retrospective study of 2,241 cases in 1989. *Acta Orthop Scand* 1994;65:432-433.
4. Owen RA, Melton LJ III, Johnson KA, Ilstrup DM, Riggs BL. Incidence of Colles' fracture in a North American community. *Am J Public Health* 1982;72:605-607.
5. Wei DH, Poolman RW, Bhandari M, Wolfe VM, Rosenwasser MP. External fixation versus internal fixation for unstable distal radius fractures: a systematic review and meta-analysis of comparative clinical trials. *J Orthop Trauma* 2012;26:386-394.
6. Darzi L. High Quality Care For All – NHS Next Stage Review Final Report. http://webarchive.nationalarchives.gov.uk/20130107105354/http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@en/documents/digitalasset/dh_085828.pdf (date last accessed 26 January 2016).
7. Fitzpatrick R, Davey C, Buxton MJ, Jones DR. Evaluating patient-based outcome measures for use in clinical trials. *Health Technol Assess* 1998;2:i-iv, 1-74.
8. Hudak PL, Amadio PC, Bombardier C. Development of an upper extremity outcome measure: the DASH (disabilities of the arm, shoulder and hand) [corrected]. The Upper Extremity Collaborative Group (UECG). *Am J Ind Med* 1996;29:602-608.
9. MacDermid JC. Development of a scale for patient rating of wrist pain and disability. *J Hand Ther* 1996;9:178-183.
10. Macey AC, Burke FD, Abbott K, et al. Outcomes of hand surgery. British Society for Surgery of the Hand. *J Hand Surg Br* 1995;20:841-855.
11. Chung KC, Pillsbury MS, Walters MR, Hayward RA. Reliability and validity testing of the Michigan Hand Outcomes Questionnaire. *J Hand Surg Am* 1998;23:575-587.
12. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992;30:473-483.
13. Levine DW, Simmons BP, Koris MJ, et al. A self-administered questionnaire for the assessment of severity of symptoms and functional status in carpal tunnel syndrome. *J Bone Joint Surg [Am]* 1993;75-A:1585-1592.
14. Meenan RF, Mason JH, Anderson JJ, Guccione AA, Kazis LE. AIMS2. The content and properties of a revised and expanded Arthritis Impact Measurement Scales Health Status Questionnaire. *Arthritis Rheum* 1992;35:1-10.
15. Smith HB. Smith hand function evaluation. *Am J Occup Ther* 1973;27:244-251.
16. Lips P, Cooper C, Agnusdei D, et al. Quality of life in patients with vertebral fractures: validation of the Quality of Life Questionnaire of the European Foundation for Osteoporosis (QUALEFFO). Working Party for Quality of Life of the European Foundation for Osteoporosis. *Osteoporos Int* 1999;10:150-160.
17. Bialocerkowski AE, Grimmer KA, Bain GI. Development of a patient-focused wrist outcome instrument. *Hand Clin* 2003;19:437-448.

18. Kori S, Miller R, Todd D. Kinesiophobia: a new view of chronic pain behavior. *Pain Management* 1990;3:35-43.
19. Rosenstiel AK, Keefe FJ. The use of coping strategies in chronic low back pain patients: relationship to patient characteristics and current adjustment. *Pain* 1983;17:33-44.
20. Altmaier EM, Russell DW, Kao CF, Lehmann TR, Weinstein JN. Role of self-efficacy in rehabilitation outcome among chronic low back pain patients. *J Couns Psychol* 1993;40:335-391.
21. Hoang-Kim A, Pegreff F, Moroni A, Ladd A. Measuring wrist and hand function: common scales and checklists. *Injury* 2011;42:253-258.
22. Changulani M, Okonkwo U, Keswani T, Kalairajah Y. Outcome evaluation measures for wrist and hand: which one to choose? *Int Orthop* 2008;32:1-6.
23. Bialocerkowski AE, Grimmer KA, Bain GI. A systematic review of the content and quality of wrist outcome instruments. *Int J Qual Health Care* 2000;12:149-157.
24. Schuind FA, Mouraux D, Robert C, et al. Functional and outcome evaluation of the hand and wrist. *Hand Clin* 2003;19:361-369.
25. Goldhahn J, Angst F, Simmen BR. What counts: outcome assessment after distal radius fractures in aged patients. *J Orthop Trauma* 2008;22:S126-S130.
26. Goldhahn J, Beaton D, Ladd A, et al. Recommendation for measuring clinical outcome in distal radius fractures: a core set of domains for standardized reporting in clinical practice and research. *Arch Orthop Trauma Surg* 2014;134:197-205.
27. 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.
28. Terwee CB, Jansma EP, Riphagen II, de Vet HC. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Qual Life Res* 2009;18:1115-1123.
29. Suk M, Hanson BP, Norvell DC, Helfet DL. The AO Handbook of Musculoskeletal Outcomes Measures and Instruments. First edition, Thieme, 2005.
30. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol* 2010;63:737-745.
31. 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.
32. MacDermid JC, Richards RS, Donner A, Bellamy N, Roth JH. Responsiveness of the short form-36, disability of the arm, shoulder, and hand questionnaire, patient-rated wrist evaluation, and physical impairment measurements in evaluating recovery after a distal radius fracture. *J Hand Surg Am* 2000;25:330-340.
33. Amadio PC, Silverstein MD, Ilstrup DM, Schleck CD, Jensen LM. Outcome after Colles fracture: the relative responsiveness of three questionnaires and physical examination measures. *J Hand Surg Am* 1996;21:781-787.
34. Lövgren A, Hellström K. Reliability and validity of measurement and associations between disability and behavioural factors in patients with Colles' fracture. *Physiother Theory Pract* 2012;28:188-197.
35. Wilcke MT, Abbaszadegan H, Adolphson PY. Evaluation of a Swedish version of the patient-rated wrist evaluation outcome questionnaire: good responsiveness, validity, and reliability, in 99 patients recovering from a fracture of the distal radius. *Scand J Plast Reconstr Surg Hand Surg* 2009;43:94-101.
36. Westphal T, Piatek S, Schubert S, Schuschke T, Winckler S. Reliability and validity of the upper limb DASH questionnaire in patients with distal radius fractures. *Z Orthop Ihre Grenzgeb* 2002;140:447-451. (In German).
37. Gabl M, Krappinger D, Arora R, et al. Acceptance of patient-related evaluation of wrist function following distal radius fracture (DRF). *Handchir Mikrochir Plast Chir* 2007;39:68-72. (In German).
38. Hemelaers L, Angst F, Drerup S, Simmen BR, Wood-Dauphinee S. Reliability and validity of the German version of "the Patient-rated Wrist Evaluation (PRWE)" as an outcome measure of wrist pain and disability in patients with acute distal radius fractures. *J Hand Ther* 2008;21:366-376.
39. MacDermid JC, Turgeon T, Richards RS, Beadle M, Roth JH. Patient rating of wrist pain and disability: a reliable and valid measurement tool. *J Orthop Trauma* 1998;12:577-586.
40. Mehta SP, Mhatre B, MacDermid JC, Mehta A. Cross-cultural adaptation and psychometric testing of the Hindi version of the patient-rated wrist evaluation. *J Hand Ther* 2012;25:65-77.
41. Kim JK, Kang JS. Evaluation of the Korean version of the patient-rated wrist evaluation. *J Hand Ther* 2013;26:238-243.
42. Schönemann JO, Hansen TB, Søballe K. Translation and validation of the Danish version of the Patient Rated Wrist Evaluation questionnaire. *J Plast Surg Hand Surg* 2013;47:489-492.
43. Walenkamp MM, de Muinck Keizer RJ, Goslings JC, et al. The Minimum Clinically Important Difference of the Patient-rated Wrist Evaluation Score for Patients With Distal Radius Fractures. *Clin Orthop Relat Res* 2015;473:3235-3241.
44. Westphal T. Reliability and responsiveness of the German version of the Disabilities of the Arm, Shoulder and Hand questionnaire (DASH). *Unfallchirurg* 2007;110:548-552. (In German.)
45. Kotsis SV, Lau FH, Chung KC. Responsiveness of the Michigan Hand Outcomes Questionnaire and physical measurements in outcome studies of distal radius fracture treatment. *J Hand Surg Am* 2007;32:84-90.
46. Shauer MJ, Chung KC. The minimal clinically important difference of the Michigan hand outcomes questionnaire. *J Hand Surg Am* 2009;34:509-514.
47. Waljee JF, Kim HM, Burns PB, Chung KC. Development of a brief, 12-item version of the Michigan Hand Questionnaire. *Plast Reconstr Surg* 2011;128:208-220.
48. Forward DP, Sithole JS, Davis TR. The internal consistency and validity of the Patient Evaluation Measure for outcomes assessment in distal radius fractures. *J Hand Surg Eur Vol* 2007;32:262-267.
49. Bialocerkowski AE, Grimmer KA, Bain GI. Validity of the patient-focused wrist outcome instrument: do impairments represent functional ability? *Hand Clin* 2003;19:449-455.
50. Goldhahn J, Shisha T, Macdermid JC, Goldhahn S. Multilingual cross-cultural adaptation of the patient-rated wrist evaluation (PRWE) into Czech, French, Hungarian, Italian, Portuguese (Brazil), Russian and Ukrainian. *Arch Orthop Trauma Surg* 2013;133:589-593.
51. Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR. Publication bias in clinical research. *Lancet* 1991;337:867-72.
52. Green A, Liles C, Rushton A, Kyte DG. Measurement properties of patient-reported outcome measures (PROMS) in Patellofemoral Pain Syndrome: a systematic review. *Man Ther* 2014;19:517-526.
53. Grevnerts HT, Terwee CB, Kvist J. The measurement properties of the IKDC-subjective knee form. *Knee Surg Sports Traumatol Arthrosc* 2015;23:3698-3706.
54. Kroman SL, Roos EM, Bennell KL, Hinman RS, Dobson F. Measurement properties of performance-based outcome measures to assess physical function in young and middle-aged people known to be at high risk of hip and/or knee osteoarthritis: a systematic review. *Osteoarthritis Cartilage* 2014;22:26-39.

Funding Statement

- None declared.

Author Contributions

- Y. V. Kleinlugtenbelt: Design of study; Acquisition, analysis and interpretation of data; Writing and revision of manuscript.
- R. W. Nienhuis: Design of study; Acquisition, analysis and interpretation of data; Revision of manuscript.
- M. Bhandari: Analysis and interpretation of data; Revision of manuscript.
- J. C. Goslings: Analysis and interpretation of data; Revision of manuscript.
- R. W. Poolman: Analysis and interpretation of data; Revision of manuscript.
- V. A. B. Scholtes: Design of study; Analysis and interpretation of data; Revision of manuscript.

ICMJE conflict of interest

- M. Bhandari reports personal fees received from Smith & Nephew, Stryker, Amgen, Zimmer, Moximed, Bioventus, Merck, Eli Lilly, Sanofi, Ferring, Conmed, as well as grants from Smith & Nephew, DePuy, Eli Lilly, Bioventus, Stryker, Zimmer, Amgen, none of which is related to this article.

© 2016 Kleinlugtenbelt et al. This is an open-access article distributed under the terms of the Creative Commons Attribution licence (CC-BY-NC), which permits unrestricted use, distribution, and reproduction in any medium, but not for commercial gain, provided the original author and source are credited.