



Universiteit
Leiden
The Netherlands

Outcome measures and their measurement properties for trapeziometacarpal osteoarthritis: a systematic literature review

Marks, M.; Schoones, J.W.; Kolling, C.; Herren, D.B.; Goldhahn, J.; Vlieland, T.P.M.V.

Citation

Marks, M., Schoones, J. W., Kolling, C., Herren, D. B., Goldhahn, J., & Vlieland, T. P. M. V. (2013). Outcome measures and their measurement properties for trapeziometacarpal osteoarthritis: a systematic literature review. *Journal Of Hand Surgery (European Volume)*, 38(8), 822-838. doi:10.1177/1753193413488301

Version: Publisher's Version

License: [Licensed under Article 25fa Copyright Act/Law \(Amendment Taverne\)](#)

Downloaded from: <https://hdl.handle.net/1887/102088>

Note: To cite this publication please use the final published version (if applicable).

Outcome measures and their measurement properties for trapeziometacarpal osteoarthritis: a systematic literature review

The Journal of Hand Surgery
(European Volume)
38E(8) 822–838
© The Author(s) 2013
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/1753193413488301
jhs.sagepub.com



M. Marks

Department of Research and Development, Schulthess Klinik, Zurich, Switzerland
Department of Orthopaedics, Leiden University Medical Center, Leiden, The Netherlands

J. W. Schoones

Walaeus Library, Leiden University Medical Center, Leiden, The Netherlands

C. Kolling

Department of Research and Development, Schulthess Klinik, Zurich, Switzerland

D. B. Herren

Department of Upper Extremities and Hand Surgery, Schulthess Klinik, Zurich, Switzerland

J. Goldhahn

Department of Research and Development, Schulthess Klinik, Zurich, Switzerland

T. P. M. Vliet Vlieland

Department of Orthopaedics, Leiden University Medical Center, Leiden, The Netherlands

Abstract

The objective was to identify all outcome measures used in studies on trapeziometacarpal osteoarthritis (TMC OA) and evaluate their measurement properties. In a two-step systematic literature review, we first identified studies including TMC OA patients and extracted all outcome measures. They were categorized according to the Outcome Measures in Rheumatology (OMERACT) core set for OA including five dimensions: pain, physical function, global assessment, imaging, and quality of life (QoL). Secondly, we retrieved articles on the measurement properties of the identified outcome measures for TMC OA patients. First, 316 articles including 101 different outcome measures were identified, addressing the OMERACT pain and function domains most frequently but under-representing QoL. Second, 12 articles investigating measurement properties of 12 outcome measures were identified. The methodological quality of these studies was poor to fair, implying that based on the literature no recommendations to use any of the outcome measures can yet be made.

Keywords

carpometacarpal joint, osteoarthritis, thumb, outcome assessment

Date received: 24th October 2012; Revised: 10th February 2013; accepted: 8th April 2013

Introduction

Numerous studies have evaluated conservative and surgical treatments for patients with trapeziometacarpal osteoarthritis (TMC OA), with both approaches generally found to be effective in reducing pain and increasing function (Mahendira and Towheed, 2009;

Martou et al., 2004; Valdes and Marik, 2010; Vermeulen et al., 2011; Wajon et al., 2009). Several specific sets

Corresponding author:

Miriam Marks, Schulthess Klinik, Department of Research and Development, Lengghalde 2, Zurich, 8008, Switzerland.
Email: Miriam.Marks@kws.ch

of outcome measures, known as core sets, are considered relevant to the best way of measuring treatment outcomes for TMC OA. Angst et al. (2005) proposed a core set to assess outcomes after resection interposition arthroplasty of the TMC joint; this consisted of the Short Form 36 (SF-36), Disabilities of the Arm, Shoulder and Hand (DASH) questionnaire, or Patient-Rated Wrist Evaluation (PRWE), and a customized form including assessment of range of motion (ROM), strength, and other clinical tests. Although showing good construct validity in this particular study, the reliability of the customized form and responsiveness of the whole set have not been investigated.

Three other core sets are available; they do not recommend specific outcome measures but rather areas that are relevant for patients with conditions affecting the hand. Based on the International Classification of Functioning, Disability and Health (ICF), a comprehensive and brief core set have been developed and validated to assess patients with any hand condition (Kus et al., 2011; 2012, Rudolf et al., 2010; 2012). These two detailed and complex core sets are known mainly to hand therapists and are not widely implemented in clinical practice. A simpler, more general core set of OA outcome measures (hip, knee, hand) was developed at the Outcome Measures in Rheumatology (OMERACT) III conference (Bellamy et al., 1997; Brooks and Hochberg, 2001). It is intended to serve as an international standard for clinical trials. This set contains the domains 'pain', 'physical function', 'patient's global assessment', 'joint imaging', and 'quality of life (QoL)', but, like the ICF core set, it does not comprise specific outcome measures.

In research and daily practice, decisions for treatments are made, amongst others, based on the results of health status questionnaires. Before such an instrument may be implemented, its measurement properties, such as reliability, validity, and responsiveness, should be assessed and considered adequate for the target population. It is important to use reliable and valid outcome measures in order to avoid biased results and conclusions (Mokkink et al., 2010a). Quality criteria for evaluating measurement properties of health status questionnaires have been introduced and are widely accepted (Lohr et al., 1996; Mokkink et al., 2010b; Scientific Advisory Committee of the Medical Outcomes Trust, 2002; Terwee et al., 2007; Valderas et al., 2008). However, these criteria do consider the outcome measure itself but not the methodological quality of the study. To evaluate whether a study on a specific outcome measure is of good methodological quality, the COnsensus-based

Standards for the selection of health status Measurement INstruments (COSMIN) checklist has been recently developed (Terwee et al., 2012).

Given the high prevalence of TMC OA (Kalichman and Hernandez-Molina, 2010; Wilder et al., 2006) and the many available treatment methods, a standardized assessment is essential for comparing the interventions and providing evidence of best practice. So far, it remains unclear to what extent researchers are using valid and reliable assessment tools in TMC OA studies and whether these meet the recommendations of the core sets mentioned previously. Furthermore, the methodological quality of studies investigating measurement properties of outcome measures for hand patients has not been investigated yet.

In order to identify suitable outcome measures and make recommendations for outcome measures to be used for patients with TMC OA, our objectives were to (1) identify all subjective and objective outcome measures used in clinical trials of conservative and surgical treatments of TMC OA; (2) relate them to the OMERACT core set; and (3) evaluate the measurement properties of standardized outcome measures employed in patients with TMC OA as well as the methodological quality of these studies.

Methods

We performed this systematic literature review in accordance with the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) statement for developing study protocols and reporting systematic reviews (Liberati et al., 2009; Moher et al., 2009). The review protocol was registered in the Netherlands National Trial Register (no. 2602).

Step 1: Literature search for TMC OA studies

An experienced librarian performed the search for articles published up to November 2010 in the following databases: PubMed, EMBASE, Web of Science, Cochrane Library, CINAHL, Academic Search Premier, ScienceDirect, and PEDro. The following search strategy was applied to PubMed and optimized for the other databases: ("trapeziometacarpal osteoarthritis" OR "trapezio metacarpal osteoarthritis" OR "carpo metacarpal osteoarthritis" OR "carpo metacarpal osteoarthritis" OR "thumb osteoarthritis" OR ([Osteoarthritis OR Osteoarthroses OR Osteoarthritis OR Osteoarthritis OR Osteoarthritis OR Osteoarthritis OR Osteoarthritis OR Osteoarthritis OR "Degenerative Arthritis" OR "Degenerative Arthritides" OR Arthrosis[tw] OR Arthroses) AND

[Carpometacarpal Joints OR Carpometacarpal Joint OR carpometacarpal OR trapeziometacarpal OR thumb OR thumbs OR "thumb base" OR carpometacarpal* OR ((Metacarpus OR Metacarpal) AND (Carpal OR Carpus OR Carpo OR Carpi)) OR ((Trapezium OR Trapezoid OR Trapezium OR Trapezial) AND (Metacarpus OR Metacarpal OR Carpal OR Carpus OR Carpo OR Carpi))]. As language restriction is unreliable or not possible in all databases, we conducted the search without any such restrictions.

Inclusion criteria for the review were (a) clinical study involving a minimum of 10 people with TMC OA who had received any conservative or surgical treatment for TMC OA; (b) study designs including all randomized controlled trials and observational (prospective or retrospective) studies; (c) studies in which the effectiveness of the treatment was evaluated with at least one outcome measure; (d) the paper was written in English or German.

Exclusion criteria were (a) studies investigating patients with generalized OA; (b) studies in which the results of patients with TMC OA could not be separated from those of patients with other conditions; (c) reviews, case reports, post-mortem and veterinary studies, and conference abstracts not published as full journal articles, because they lacked full information about the study design; and (d) studies not in English or German, as we had no reliable translators.

Two independent reviewers reviewed the titles and abstracts that had been identified. The full texts of the selected abstracts were retrieved and again analyzed independently by two of the authors. Consensus on inclusion of the studies was reached by discussion.

We checked the references of the included articles to find other suitable papers and subjected them to a similar selection process.

Data for the following variables were extracted using a predefined form: authors, publication year, number of patients, patient demographics, intervention, and follow-up period, as well as all objective and subjective outcome measures used in the studies. The level of evidence was determined using the slightly modified rating scheme described by Wright et al. (2003).

Some authors analyzed the same study population more than once and presented their findings in several publications. These articles were analyzed as individual studies, in case inadequate descriptions of the study populations made it impossible to identify the overlap.

We classified all concepts included in the outcome measures, such as pain or strength, according

to the five domains of the OMERACT core set ('pain', 'physical function', 'patient's global assessment', 'joint imaging', 'QoL') (Bellamy et al., 1997). We chose this core set as the reference tool because it is simpler and better known to hand surgeons than the complex ICF concept. For the purpose of this study, the domain 'function' included isolated functions of the hand (such as extending the thumb) and activities of daily living (ADLs) making use of the hand; 'global assessment' was defined as an overall assessment of the hand condition, including treatment satisfaction, symptom improvement, and disease activity. 'QoL' was defined as a multidimensional appraisal of various aspects of health, including pain and function. 'Imaging' included all techniques such as radiography or magnetic resonance imaging. Given that some outcome measures cover more than one concept, each item, element, or dimension of a combined outcome measure was analyzed separately to assign it to several corresponding OMERACT domains.

Step 2: Measurement properties

In the period up to April 2012, we performed a second literature search on the measurement properties of the identified outcome measures for TMC OA in the databases mentioned previously. We applied the same strategy as in step 1 adding the following terms on measurement properties:

... AND (Psychometrics OR Psychometric OR Psychometr* OR "psychological variable" OR "psychological variables" OR Validity OR valid OR validated OR validation OR Validities OR "Validation Studies"[Publication Type] OR valid* OR Reliability OR Reliable OR Unreliability OR Unreliable OR Responsiveness OR Unresponsiv* OR Irresponsiv* OR Responsive* OR "Reproducibility of Results"[Mesh] OR Reproducibility OR Reproducible OR Irreproducib* OR Reliabilities). A cited reference search for the target articles was also carried out.

We included studies with a population of at least 50% of the patients suffering TMC OA or analyzing patients with TMC OA as a subgroup, and evaluating any measurement property of an outcome measure revealed in step 1, regardless of whether the investigation of measurement properties was the primary objective of the study or only mentioned tangentially. Studies investigating patients with hand OA, for example, where the proportion of TMC OA patients was less than 50%, were excluded.

The following eight measurement properties of the outcome measures were rated according to the criteria developed by Terwee et al. (2007), which we slightly

modified for our purpose (see definitions in Appendix 1): internal consistency, content validity, criterion validity, construct validity, reproducibility (agreement and reliability), responsiveness, floor or ceiling effects, and interpretability. Two reviewers independently extracted all these data and results were graded as positive (+), doubtful (?), or poor (-). As several studies investigated the same tool, the different studies were synthesized using the rating achieved by most of the articles.

The methodological quality of the articles reporting on the measurement properties of outcome measures was rated on a 4-point scale according to the COSMIN checklist (Terwee et al., 2012). This checklist is used to assess whether a study on a specific outcome measure tool meets the standards for good methodological quality. A score is calculated for each of nine standards (COSMIN boxes A-I), which somewhat differ from those criteria of Terwee (2007): 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. There are two additional boxes given: the generalizability and interpretability box. The corresponding 15 items are intended to be used as data extraction forms to extract all data on study characteristics and interpretability issues (e.g., norm scores, floor/ceiling effects, and relevancy for subgroups) (Terwee et al., 2012). An assessments of the statistical methods used in articles based on the Item Response Theory (IRT) (box general requirements for studies that applied IRT models) was not performed, as this procedure was not used in any of the included studies. Each standard (box) included various items (number ranging from 5 to 18 per box). An overall quality score for that standard was obtained by taking the worst rating of any item (worst score counts principle). The resulting rating could be excellent, good, fair, or poor (Terwee et al., 2012). There is no formal interpretation of how to combine the measurement property scores (Terwee et al.'s checklist) and methodological quality scores of studies according to the COSMIN checklist. The COSMIN group stated that the quality of an instrument under investigation remains unclear if the methodological quality of a study is inadequate (Terwee et al., 2012). For that reason, in the present study, we considered the measurement properties of a tool to be equivocal if the methodological quality of the related studies was rated as poor, irrespective of its rating on the Terwee scale.

Results

Step 1: Literature search for TMC OA studies

Our initial search identified 2979 articles. After removing duplicates, checking references, and the two-phase review process, we finally included 316 articles (Figure 1, references in Appendix 2) investigating 13 231 patients (Table 1). Forty-five articles from 17 different research groups reported on patients who had also been subjects in other studies included in our review. Four articles reported on 273 patients affected by hand OA, but the precise number of patients with TMC OA could not be determined (Haugen et al., 2010; Kvien et al., 2008; Rogers and Wilder, 2007; 2009). Different surgical procedures were investigated in 268 articles, while conservative treatments were studied in 66 papers. The methodological quality of most of the articles was low: 244 were level IV studies.

In total, we identified 101 different outcome measures, not counting 22 ways to examine radiographs and the self-developed instruments that were excluded.

These 101 outcome measures addressed the OMERACT domain 'pain' in 298 articles, 'physical function' in 303, 'global assessment' in 187, 'imaging' in 213, and 'QoL' in 13 (Table 2). A visual analogue scale (VAS) was most often applied ($n = 93$) in the domain 'pain'. 'Physical function' most frequently included measurement of muscle strength and ROM. Grip strength ($n = 218$) was the most commonly assessed measure of strength, often using a dynamometer ($n = 122$). Thumb ROM was most often based on abduction ($n = 179$), in most cases not stating the method used to measure it ($n = 114$), but sometimes mentioning use of a goniometer ($n = 26$).

'Global assessment' was done primarily by evaluating treatment satisfaction ($n = 160$) using nine different tools. 'Imaging' consisted mainly of rating the stage of OA on the radiographs ($n = 160$), most frequently using the Eaton classification ($n = 132$). The Colville questionnaire was used to evaluate 'QoL' in five of the 13 articles investigating this dimension.

Twenty-one different standardized questionnaires were used; the DASH was the most common, having been applied in 46 articles.

Step 2: Measurement properties

The second literature search yielded 538 articles, of which we included 12 (Angst et al., 2005; 2009; Citron et al., 2007; De Smet, 2004; Dela Rosa et al., 2004;

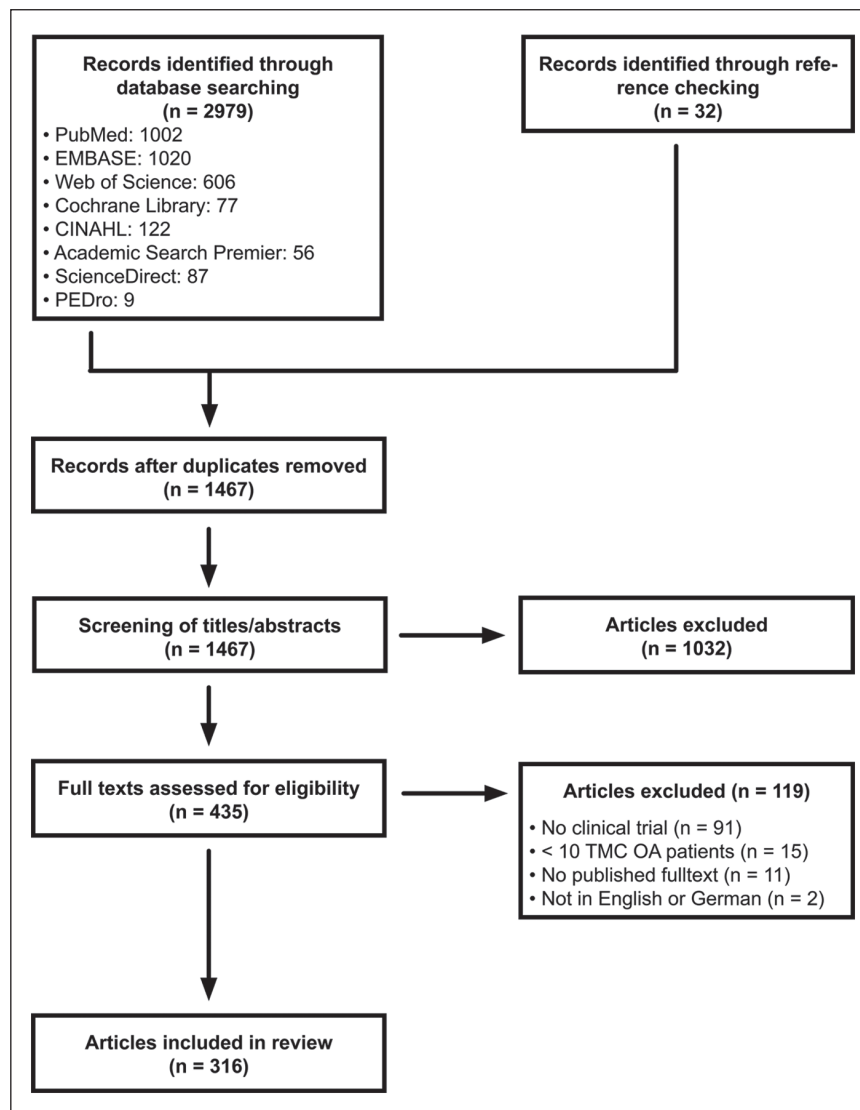


Figure 1. Study selection process for step 1.

Hansen et al., 2012; John et al., 2008; Kubik III and Lubahn, 2002; MacDermid et al., 2007; Merritt et al., 2010; Niekel et al., 2009; Spaans et al., 2011) in the final analysis (Figure 2, Table 3).

These articles examined the measurement properties of 12 outcome measures specifically in patients with TMC OA (Table 4). The DASH and the PRWE were the ones most extensively studied. None of the studies examined all eight measurement properties. Positive ratings (+) were seen for the DASH (Angst et al., 2005; 2009; Citron et al., 2007; De Smet, 2004; John et al., 2008; MacDermid et al., 2007; Niekel et al., 2009), quickDASH (Angst et al., 2009; Niekel et al., 2009), Australian/Canadian Osteoarthritis Hand Index (AUSCAN) (MacDermid et al., 2007), and Nelson Score (Citron et al., 2007). In

contrast, the Eaton classification (Dela Rosa et al., 2004; Hansen et al., 2012; Kubik III and Lubahn, 2002; Spaans et al., 2011), CMC grind test (Merritt et al., 2010), and Hand Functional Index of the Keitel Functional Test (HFI/KFT) (Angst et al., 2005) rated poorly. Ratings for the PRWE (Angst et al., 2005; John et al., 2008; MacDermid et al., 2007) and SF-36 (Angst et al., 2005; 2009; John et al., 2008; MacDermid et al., 2007) were equivocal.

The methodological quality of these articles, rated according to the COSMIN checklist was generally fair to poor and most of the measurement properties have not been investigated (Figure 3, Table 5). The positive results of the DASH were weakened by the poor methodological quality of the studies investigating its responsiveness (De Smet, 2004), while the overall

Table 1. Characteristics of the 316 clinical studies on TMC OA. Due to inadequate descriptions, not all variables could be extracted from all studies.

	Sum of all studies [% of all articles]	Median (range) per study
Year		2000 (1968–2010)
Patients^a	13 231	32 (10–315)
Females^b (% of population)	8855 (83.2)	26 (0–162)
Males^b (% of population)	1784 (16.8)	5 (0–38)
Hands^c	12 521	34 (0–315)
Age^d (years)		59.1 (33.7–74.6)
Follow up^e (years)		2.9 (0.04–16.4)
Level of evidence (%)^f		4 (1–4)
Level I	33 (10)	
Level II	13 (4)	
Level III	26 (8)	
Level IV	244 (77)	
Intervention^g		
Implant arthroplasty	92	
Trapeziectomy + ligament reconstruction + tendon interposition	67	
Trapeziectomy + tendon interposition	49	
Trapeziectomy	36	
Arthrodesis	33	
Injection	28	
Splint	16	
Trapeziectomy + interposition with various material	15	
Various surgical interventions ^h	14	
Trapeziectomy + ligament reconstruction	11	
Various conservative treatments ⁱ	8	
Osteotomy	8	
Physical/occupational therapy	5	
Drugs	5	
Unspecified conservative treatments	4	

^aTaken from 315 articles.^bTaken from 259 articles.^cTaken from 270 articles.^dTaken from 273 articles.^eTaken from 287 articles.^fDue to rounding errors, the sum of the percentages may be < 100%.^gMore than one intervention per study possible.^hIncluding unspecified surgical interventions, different surgical interventions in one study group, tendon interposition without trapeziectomy, debridement, synovectomy, or denervation.ⁱIncluding laser therapy, iontophoresis, radiation therapy, leech therapy, nettle sting, acupuncture, phonophoresis.

quality of the study considering the Nelson score was also rated as poor (Citron et al., 2007).

Discussion

In our review of the outcome measures used in TMC OA studies, we identified 316 papers. We found a wide variety of outcome measures, with pain and function being the most frequent and QoL underrepresented. Studies rarely examined the measurement properties of outcome measures specifically for patients with TMC OA, and the methodological quality of those that

did so was fair, so that no recommendations for the use of any outcome measure can be made.

The heterogeneity of the outcome measures employed raises serious issues about the statistical comparison of different interventions, as shown in a recent systematic review of the surgical management of TMC OA (Vermeulen et al., 2011). This concerns not only studies on patients with TMC OA, but also studies on hand OA, where many different outcome measures have also been used (Michon et al., 2011). The finding that numerous tools (some self-developed) were used to assess the effectiveness of treatment highlights

Table 2. Concepts and outcome measures used in 316 articles about TMC OA categorized according to the OMERACT core set. The OMERACT domain is given in capital letters. Furthermore, outcome measures are arranged according to whether they are specific for the hand/upper extremity or if they are generic outcome measures.

OMERACT domain and outcome measure	Articles (n)
PAIN	298^a
Hand specific	
Visual Analogue Scale (VAS)	93
Likert scale(s)	48
Joint tenderness	23
Carpometacarpal grind test	14
Alnot classification	5
Self-developed questionnaire for hand pain	111
Generic	
Intake of analgesics	27
McGill Pain Questionnaire	1
PHYSICAL FUNCTION	303^a
Hand specific	
Strength	267
Range of motion	223
Stability	42
Dexterity	30
Sensibility	25
Subjective hand function	24
Stiffness	19
Wound healing	9
Self-developed function tests	9
Pegboard tests	8
Functional Index of Hand OA (FIHOA)/Dreiser index	5
Jebsen-Taylor test	4
Muscle outline	2
Cochin Scale	2
Sollerman Hand Function Test	2
Green Test	1
Upper extremity specific	
Activities of daily living (ADLs) — self-developed questionnaire	75
Activities of daily living (ADLs) — method not specified	25
Hand Functional Index (HFI) of the Keitel Functional Test (KFT)	2
Abilhand	1
Generic	
Sleep disturbance	2
Fatigue	1
GLOBAL ASSESSMENT	187^a
Treatment satisfaction	160
Subjective result	22
Self-developed questionnaire	4
Disease activity	1
IMAGING	213^a
Stage of thumb OA (radiographs)	160
Scapho-metacarpal distance	104
QUALITY OF LIFE	13^a
Colville questionnaire	5
Arthritis Impact Measurement 2 (AIMS2)	3
Medical Outcomes Study Short Form 36 (SF-36)	4
Medical Outcomes Study Short Form 12 (SF12)	1

(Continued)

Table 2. (Continued)

OMERACT domain and outcome measure	Articles (n)
PAIN + FUNCTION	291^a
Hand specific	
Australian/Canadian Osteoarthritis Hand Index (AUSCAN)	5
Patient-Rated Wrist Evaluation (PRWE)	4
Sequential Occupational Dexterity Assessment (SODA)	3
Nelson Score	1
Upper extremity specific	
Disabilities of the Arm, Shoulder and Hand (DASH) questionnaire	46
QuickDASH	3
Generic	
Health Assessment Questionnaire – Disability Index (HAQ-DI)	3
Western Ontario McMaster Universities Osteoarthritis Index (WOMAC)	1
PAIN + FUNCTION + GLOBAL	173^a
Hand specific	
Buck Gramko scale	10
Michigan Hand Questionnaire (MHQ)	4
Patient Evaluation Measure (PEM)	1
OMERACT-OARSI response index	1
OMERACT NOT ASSIGNABLE	
Hand specific	
Hand appearance	43
Crepitus	24
Thumb shortening	10
Confidence with hand use	1
Generic	
Complications	234
Return to work	69
Comfort with device	6
Laboratory results	2
Met expectations	2
Intake of hormones	2
Pain Anxiety Symptoms Scale (PASS)	1
Center for the Epidemiological Study of Depression instrument (CES-D)	1
Pain Catastrophizing Scale (PCS)	1

^aNumber of articles covering this domain. Although pain, for example, might be evaluated by more than one outcome measure, this value does not necessarily reflect the sum of the instruments given below.

the need to develop homogeneous, standardized, and validated outcome measures for patients with TMC OA in order to facilitate comparisons of patient populations and the outcomes of different surgical and non-surgical procedures.

Apart from the variety, we also found that specific aspects of outcome were not covered equally. The OMERACT core set includes the assessment of QoL as a strongly recommended module (Bellamy et al., 1997), but only few studies on TMC OA include it. Given that hand OA greatly affects QoL (Michon et al., 2011), several authors recommend using a generic outcome measure such as the SF-36 to evaluate QoL in patients

with hand disorders (Angst et al., 2005; Goldhahn et al., 2008; Maheu et al., 2006; Michon et al., 2011). The observed predominance of objective measures (such as muscle strength and ROM) performed by healthcare providers shows that many researchers still do not make the subjective patient perspective their primary focus. This implies underrepresentation of concepts such as psychological consequences, aesthetic changes, and effects on leisure activities, which are important to patients with hand OA (Stamm et al., 2009).

The measurement properties of the DASH and PRWE were the most extensively examined ones in

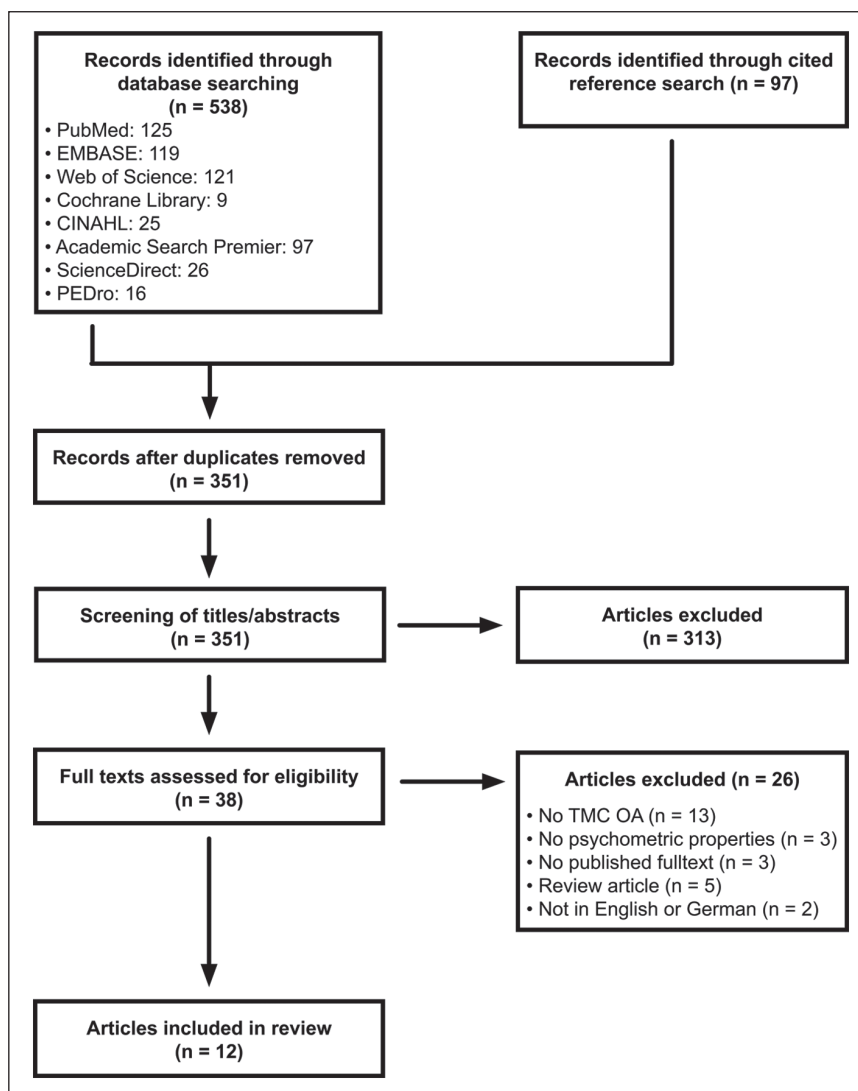


Figure 2. Study selection process for step 2.

patients with TMC OA. Overall, the DASH was rated more favourably than the PRWE, especially regarding responsiveness, and floor and ceiling effects. It should be noted, however, that the methodological quality of the studies on the responsiveness of DASH was low [Citron et al., 2007; De Smet, 2004]. If the study methodology is of poor quality, the quality of the instrument remains equivocal [Terwee et al., 2012]. Furthermore, the specificity and sensitivity of this tool in these particular patients remains questionable because the score is influenced by function/dysfunction of the elbow and shoulder [Citron et al., 2007; Goldhahn et al., 2008; MacDermid and Tottenham, 2004]. For this reason, it might be better to use a hand-specific questionnaire such as the AUSCAN or Nelson score. The AUSCAN has only been examined for construct validity in patients with

TMC OA, which does not permit any firm conclusions on its overall value in this patient group. The reliability and responsiveness of the AUSCAN were, however, found to be satisfactory for patients with general hand OA [Haugen et al., 2010; Moe et al., 2010]. Apart from its measurement properties, other characteristics of a questionnaire such as feasibility and associated costs have to be considered. While the DASH is freely available, the AUSCAN has to be purchased. The Nelson score, a questionnaire specifically designed to assess the outcome following TMC OA surgery, has so far only been applied by the developers themselves [Citron et al., 2007]. Interpretation of their findings is further hampered by the poor methodological quality of the study, including assessing only 36 patients. The Eaton classification to assess the stage of OA is the only imaging method that has

Table 3. Study characteristics of the 12 included articles about measurement properties of outcome measures for patients with TMC OA according to the COSMIN generalizability and interpretability boxes.

Study	Outcome measure	No. of patients with TMC OA	Age, years (mean, SD)	Sex (% females)	Treatment	Setting	Countries	Language	Patient selection	Response rate (%)
John et al., 2008*	PRWE, DASH, SF-36	103 (112 cases)	68 ± 9.8	83	Resection Interposition Arthroplasty	Hospital	Switzerland	German	Consecutive	72
Angst et al., 2009*	quickDASH, DASH, SF-36	103 (112 cases)	68 ± 9.8	83	Resection Interposition Arthroplasty	Hospital	Switzerland	German	Consecutive	72
Angst et al., 2005*	PRWE, DASH, SF-36, HFI/KFT	103 (112 cases)	68 ± 9.8	83	Resection Interposition Arthroplasty	Hospital	Switzerland	German	Consecutive	72
MacDermid et al., 2007	PRWE, DASH, AUSCAN, SF-36	120	65 ± 8.1	82	Resection Interposition Arthroplasty	Hospital	Canada			
De Smet, 2004	DASH	15	56 (median)	93	Surgery	Hospital	Belgium		Consecutive	
Niekel et al., 2009	DASH, quickDASH, CES-D, PCS, PASS	107			None described	Hospital	The Netherlands		Convenience	27
Citron et al., 2007	Nelson score	36			Surgery	Outpatient clinic	UK	English	Consecutive	
Dela Rosa et al., 2004	Eaton classification	30 (40 cases)	59	87	None described	Hospital	USA	n/a	Random	n/a
Kubik III and Lubahn, 2002	Eaton classification	40	60	83	None described	Hospital	USA	N/A	Random	N/A
Hansen et al., 2012	Eaton classification	43 (50 cases)	60	72	Pre-operative analysis	Hospital	Denmark	N/A		N/A
Spaans et al., 2011	Eaton classification	40 cases	60	73	Various	Hospital	The Netherlands	N/A	Convenience	N/A
Merritt et al., 2010	CMC grind test	54 (70 cases)	60 ± 13.4	85	None	Private orthopaedic clinic, occupational medicine clinic, general community	USA	N/A	Convenience	N/A

(Continued)

Table 3. (Continued)

Study	Outcome measures	Missing items	Handling of missing items	Distribution of the scores	Floor effect	Ceiling effect	Scores for relevant (sub) groups	MIC or MID
John et al., 2008*	PRWE, DASH, SF-36	PRWE: ≤ 12%		PRWE: left skewed	PRWE: 16–24%			
Angst et al., 2009*	quickDASH, DASH, SF-36							quickDASH/DASH: 4.5–11.3
Angst et al., 2005*	PRWE, DASH, SF-36, HFI/KFT	PRWE ≤ 14%; DASH ≤ 17%; SF-36 ≤ 6%; HFI/KFT = 7%		Non-parametric	PRWE ≤ 24%; DASH = 0%; SF-36 ≤ 16%; HFI/KFT = 0%	PRWE = 0%; DASH ≤ 20%; SF-36 ≤ 83%; HFI/KFT = 16%	Norm data given for SF-36 and DASH	
MacDermid et al., 2007	PRWE, DASH, AUSCAN, SF-36			SF-36: normal distribution; other instruments: non-parametric	Existent for PRWE, DASH, and AUSCAN	Existent for AUSCAN	Scores for patients with solely hand OA compared to patients with hand OA and OA at other joints	
De Smet, 2004	DASH			Non-parametric	Existent	Not existent	Scores for other hand disorders included	
Nielke et al., 2009	DASH, quickDASH, CES-D, PCS, PASS							
Citron et al., 2007	Nelson score							
Dela Rosa et al., 2004	Eaton classification	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Kubik III and Lubahn, 2002	Eaton classification	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Hansen et al., 2012	Eaton classification	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Spaans et al., 2011	Eaton classification	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Merritt et al., 2010	CMC grind test	N/A	N/A	N/A	N/A	N/A	N/A	N/A

*These articles report on the same cohort.

AUSCAN = Australian/Canadian Osteoarthritis Hand Index; CES-D: Center for the Epidemiological Study of Depression instrument; DASH = Disabilities of the Arm, Shoulder and Hand questionnaire; HFI/KFT: Hand Functional Index of the Keitel Functional Test; N/A: not applicable; PASS: Pain Anxiety Symptoms Scale; PCS: Pain Catastrophizing Scale; PRWE: Patient-Rated Wrist Evaluation; SF-36: Medical Outcomes Study Short Form 36.

Table 4. Rating of the measurement properties of outcome measures tested for patients with TMC OA according to Terwee et al. [2007]. They were graded as positive (+), doubtful (?), or poor (–). Blank boxes show that these properties have not been investigated. References are given in parentheses.

Outcome measure instrument	Internal Consistency	Content validity	Criterion validity	Construct validity	Reproducibility	Responsiveness	Floor or ceiling effect	Interpretability
DASH	+		+	+	+	+	+	?
	(Angst et al., 2009)		(Niekel et al., 2009)	(Angst et al., 2005; 2009; De Smet, 2004; John et al., 2008; MacDermid et al., 2007)		(Citron et al., 2007; De Smet, 2004)	(Angst et al., 2005)	(MacDermid et al., 2007)
AUSCAN (MacDermid et al., 2007)				+				?
PRWE	+			+	+	–	–	?
	(John et al., 2008)			(Angst et al., 2005; John et al., 2008)	(John et al., 2008)	(John et al., 2008)	(Angst et al., 2005; John et al., 2008)	(MacDermid et al., 2007)
Nelson Score (Citron et al., 2007)	+	?		?	?	+		
SF-36	+			?			–	
	(Angst et al., 2009)			(Angst et al., 2005; 2009; John et al., 2008; MacDermid et al., 2007)			(Angst et al., 2005)	
Eaton classification (stage of OA) (Dela Rosa et al., 2004; Hansen et al., 2012; Kubik III and Lubahn, 2002; Spaans et al., 2011)					–			
Carpometacarpal grind test (Merritt et al., 2010)			?		–			
quickDASH	+		+	?				
	(Angst et al., 2009)		(Angst et al., 2009; Niekel et al., 2009)	(Angst et al., 2009; Niekel et al., 2009)				
CES-D (Niekel et al., 2009)				?				
PCS (Niekel et al., 2009)				?				
PASS (Niekel et al., 2009)				?				
HFI/KFT (Angst et al., 2005)				–			–	

AUSCAN = Australian/Canadian Osteoarthritis Hand Index; CES-D = Center for the Epidemiological Study of Depression instrument; DASH = Disabilities of the Arm, Shoulder and Hand questionnaire; HFI/KFT = Hand Functional Index of the Keitel Functional Test; PASS = Pain Anxiety Symptoms Scale; PCS = Pain Catastrophizing Scale; PRWE = Patient-Rated Wrist Evaluation; SF-36 = Medical Outcomes Study Short Form 36.

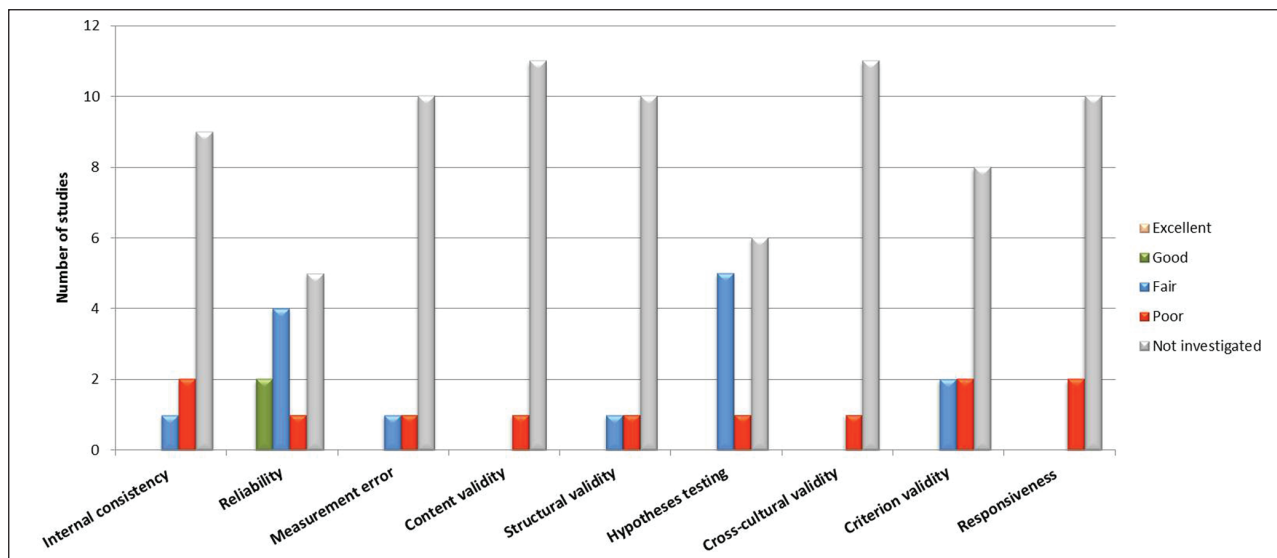


Figure 3. Distribution of the methodological quality of 12 studies about measurement properties rated with the COSMIN checklist.

been studied for reliability in patients with TMC OA. Although its reliability remains questionable, it seems to be the best method of staging currently available (Spaans et al., 2011). The patient's global assessment was done primarily by evaluating patient satisfaction. Researchers used several instruments, such as a VAS, Likert scale is, and different questionnaires, all of which still have to be tested for their measurement properties in patients with TMC OA. To date, there is no validated instrument available in hand surgery to measure patient satisfaction, which might be due to the numerous health-related, personal, and environmental factors influencing patient satisfaction (Marks et al., 2011). The present review yielded equivocal ratings regarding construct validity and floor effect for the SF-36 with respect to QoL, and its responsiveness has not been investigated for TMC OA patients. Although other researchers have found a relatively low sensitivity to change in patients with carpal tunnel syndrome (Gay et al., 2003; Rosales et al., 2009; Uchiyama et al., 2007) and distal radius fractures (Imaeda et al., 2010; MacDermid et al., 2000), a generic instrument to measure QoL is recommended because it allows the comparison between different conditions and patient populations (Guyatt, 1995).

Assessing the methodological quality of studies is an important point in systematic reviews. However, there are no uniform guidelines on how to assess the methodological quality for different types of studies. For randomized controlled trials (RCTs), the Cochrane collaboration recommends its risk of bias tool

(Higgins and Green, 2011). For observational studies, there are various checklists and scores available, but none of these can be recommended to be used as a gold-standard (Sanderson et al., 2007). Other common checklists, such as the CONSORT (Schulz et al., 2010), PRISMA (Liberati et al., 2009; Moher et al., 2009), and Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) (Vandenbroucke et al., 2007) statements, are not intended to serve as quality appraisal tools but to guide authors when reporting RCTs, systematic reviews and observational studies,, respectively. For grading the methodological quality of studies investigating measurement properties, the COSMIN checklist (Terwee et al., 2012) is the only available tool, so far.

Our review has certain limitations. As only English and German articles have been included, some studies published in other languages might have been missed. Additionally, many articles lacked information on the study population and methods, making it impossible to determine actual overlap among studies and calculate the exact numbers of patients investigated. Furthermore, the low methodological quality of all the studies, assessed by the COSMIN checklist, prohibits recommendations. The scoring of this tool is rather rigid, giving the overall rating of a specific measurement property as poor even if only one item is scored as such. For each measurement property, the number of missing items and their handling has to be scored. Though this information is lacking in most of the studies, this leads to an overall fair rating,

Table 5. Methodological quality of studies investigating outcome measure instruments in patients with TMC OA. In accordance with the COSMIN checklist, the method of investigating each measurement property was rated excellent, good, fair, or poor.

Author	Outcome measure	Internal consistency	Reliability	Measurement error	Content validity	Structural validity	Hypotheses testing	Cross-cultural validity	Criterion validity	Responsiveness
John et al., 2008*	PRWE, DASH, SF-36	Poor	Fair	Fair			Fair	Poor	Poor	
Angst et al., 2009*	quickDASH, DASH, SF-36	Fair					Fair		Fair	
Angst et al. 2005*	PRWE, DASH, SF-36, HFI/KFT					Poor	Fair			
MacDermid et al., 2007	PRWE, DASH, AUSCAN, SF-36					Fair	Fair			
De Smet, 2004	DASH									Poor
Nielke et al., 2009	DASH, quickDASH, CES-D, PCS, PASS						Fair		Poor	
Citron et al., 2007	Nelson score, DASH									Poor
Dela Rosa et al., 2004	Eaton classification	Poor	Poor	Poor	Poor		Poor			
Kubik III and Lubahn, 2002	Eaton classification		Fair							
Hansen et al., 2012	Eaton classification		Good							
Spanns et al., 2011	Eaton classification		Fair							
Merritt et al., 2010	CMC grind test		Fair						Fair	

*These articles report on the same cohort.

AUSCAN = Australian/Canadian Osteoarthritis Hand Index; CES-D: Center for the Epidemiological Study of Depression instrument; DASH = Disabilities of the Arm, Shoulder and Hand questionnaire; HFI/KFT: Hand Functional Index of the Keitel Functional Test; PASS: Pain Anxiety Symptoms Scale; PCS: Pain Catastrophizing Scale; PRWE: Patient-Rated Wrist Evaluation; SF-36: Medical Outcomes Study Short Form 36.

although the study achieved better ratings regarding the other items of that property. For this reason, the methodological quality of the articles might have been underestimated. Another limitation of the study is that we used the Wright classification for rating the levels of evidence. Following our rating, a revised classification for evidence-based medicine was published (Howick et al., 2011). As the primary purpose of our publication was not to report the evidence levels of studies on TMC OA but rather to focus on measurement instruments, it was decided not to repeat the classification.

Based on the results of the present study, no recommendation for a particular outcome measure can be made. A combination of hand-specific questionnaires, which are most suitable for detecting changes in patients with TMC OA, general health status questionnaires, and clinical data are suggested. However, more research on the psychometric properties of outcome measures in methodologically sound studies is needed before we can make any firm recommendations about the use of specific tools.

Acknowledgements

We thank Stefanie Hensler, Martina Wehrli, and Carina Muoth for assistance in reviewing the articles, Prof Dr Nelissen for his scientific support, and Dr Meryl Clarke for her assistance in preparing the manuscript.

Conflict of interests

None declared.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

References

- Angst F, John M, Goldhahn J et al. Comprehensive assessment of clinical outcome and quality of life after resection interposition arthroplasty of the thumb saddle joint. *Arthritis Rheum.* 2005, 53: 205–13.
- Angst F, Goldhahn J, Drerup S, Flury M, Schwyzer HK, Simmen BR. How sharp is the short QuickDASH? A refined content and validity analysis of the short form of the Disabilities of the Shoulder, Arm and Hand questionnaire in the strata of symptoms and function and specific joint conditions. *Qual Life Res.* 2009, 18: 1043–51.
- Bellamy N, Kirwan J, Boers M et al. Recommendations for a core set of outcome measures for future phase III clinical trials in knee, hip, and hand osteoarthritis. Consensus development at OMERACT III. *J Rheumatol.* 1997, 24: 799–802.
- Brooks P, Hochberg M. Outcome measures and classification criteria for the rheumatic diseases. A compilation of data from OMERACT (Outcome Measures for Arthritis Clinical Trials), ILAR (International League of Associations for Rheumatology), regional leagues and other groups. *Rheumatology (Oxford).* 2001, 40: 896–906.
- Citron N, Hulme CE, Wardle N. A self-administered questionnaire for basal osteoarthritis of the thumb. *J Hand Surg Eur.* 2007, 32: 524–8.
- De Smet L. Responsiveness of the DASH score in surgically treated basal joint arthritis of the thumb: preliminary results. *Clin Rheumatol.* 2004, 23: 223–4.
- Dela Rosa TL, Vance MC, Stern PJ. Radiographic optimization of the Eaton classification. *J Hand Surg Br.* 2004, 29: 173–7.
- Gay RE, Amadio PC, Johnson JC. Comparative responsiveness of the Disabilities of the Arm, Shoulder, and Hand, the carpal tunnel questionnaire, and the SF-36 to clinical change after carpal tunnel release. *J Hand Surg Am.* 2003, 28: 250–4.
- Goldhahn J, Angst F, Simmen BR. What counts: outcome assessment after distal radius fractures in aged patients. *J Orthop Trauma.* 2008, 22: 126–30.
- Guyatt GH. A taxonomy of health status instruments. *J Rheumatol.* 1995, 22: 1188–90.
- Hansen TB, Sorensen OG, Kirkeby L, Homilius M, Amstrup AL. Computed tomography improves intra-observer reliability, but not the inter-observer reliability of the Eaton-Glickel classification. *J Hand Surg Eur.* 2012, 38: 187–91.
- Haugen IK, Slatkowsky-Christensen B, Lessem J, Kvien TK. The responsiveness of joint counts, patient-reported measures and proposed composite scores in hand osteoarthritis: analyses from a placebo-controlled trial. *Ann Rheum Dis.* 2010, 69: 1436–40.
- Higgins J, Green S. *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0* [updated March 2011]. The Cochrane Collaboration, 2011.
- Howick J, Chalmers I, Glasziou P. *The 2011 Oxford CEBM Levels of Evidence (Introductory Document).* Oxford Centre for Evidence-Based Medicine, 2011. <http://www.cebm.net/index.aspx?o=5653> [accessed 18 April 2013].
- Imaeda T, Uchiyama S, Wada T et al. Reliability, validity, and responsiveness of the Japanese version of the Patient-Rated Wrist Evaluation. *J Orthop Sci.* 2010, 15: 509–17.
- John M, Angst F, Awiszus F, Pap G, Macdermid JC, Simmen BR. The Patient-Rated Wrist Evaluation (PRWE): cross-cultural adaptation into German and evaluation of its psychometric properties. *Clin Exp Rheumatol.* 2008, 26: 1047–58.
- Kalichman L, Hernandez-Molina G. Hand osteoarthritis: an epidemiological perspective. *Semin Arthritis Rheum.* 2010, 39: 465–76.
- Kubik NJ III, Lubahn JD. Intrarater and interrater reliability of the Eaton classification of basal joint arthritis. *J Hand Surg Am.* 2002, 27: 882–5.
- Kus S, Dereskewitz C, Schwab M, Eisenschenk A, Steen M, Rudolf K-D. Validation of the Comprehensive International Classification of Functioning, Disability and

- Health (ICF) core set for hand conditions. *Hand Therapy*. 2011, 16: 58–66.
- Kus S, Oberhauser C, Cieza A. Validation of the brief International Classification of Functioning, Disability, and Health (ICF) core set for hand conditions. *J Hand Ther*. 2012, 25: 274–87.
- Kvien TK, Fjeld E, Slatkowsky-Christensen B et al. Efficacy and safety of a novel synergistic drug candidate, CRx-102, in hand osteoarthritis. *Ann Rheum Dis*. 2008, 67: 942–8.
- Liberati A, Altman DG, Tetzlaff J et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ*. 2009, 339: b2700.
- Lohr KN, Aaronson NK, Alonso J et al. Evaluating quality-of-life and health status instruments: development of scientific review criteria. *Clin Ther*. 1996, 18: 979–92.
- 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: 330–40.
- MacDermid JC, Tottenham V. Responsiveness of the Disability of the Arm, Shoulder, and Hand (DASH) and Patient-Rated Wrist/Hand Evaluation (PRWHE) in evaluating change after hand therapy. *J Hand Ther*. 2004, 17: 18–23.
- MacDermid JC, Wessel J, Humphrey R, Ross D, Roth JH. Validity of self-report measures of pain and disability for persons who have undergone arthroplasty for osteoarthritis of the carpometacarpal joint of the hand. *Osteoarthritis Cartilage*. 2007, 15: 524–30.
- Mahendira D, Towheed TE. Systematic review of non-surgical therapies for osteoarthritis of the hand: an update. *Osteoarthritis Cartilage*. 2009, 17: 1263–8.
- Maheu E, Altman RD, Bloch DA et al. Design and conduct of clinical trials in patients with osteoarthritis of the hand: recommendations from a task force of the Osteoarthritis Research Society International. *Osteoarthritis Cartilage*. 2006, 14: 303–22.
- Marks M, Herren DB, Vliet Vlieland TP, Simmen BR, Angst F, Goldhahn J. Determinants of patient satisfaction after orthopedic interventions to the hand: a review of the literature. *J Hand Ther*. 2011, 24: 303–12.
- Martou G, Veltri K, Thoma A. Surgical treatment of osteoarthritis of the carpometacarpal joint of the thumb: a systematic review. *Plast Reconstr Surg*. 2004, 114: 421–32.
- Merritt MM, Roddey TS, Costello C, Olson S. Diagnostic value of clinical grind test for carpometacarpal osteoarthritis of the thumb. *J Hand Ther*. 2010, 23: 261–7.
- Michon M, Maheu E, Berenbaum F. Assessing health-related quality of life in hand osteoarthritis: a literature review. *Ann Rheum Dis*. 2011, 70: 921–8.
- Moe RH, Garratt A, Slatkowsky-Christensen B et al. Concurrent evaluation of data quality, reliability and validity of the Australian/Canadian Osteoarthritis Hand Index and the Functional Index for Hand Osteoarthritis. *Rheumatology (Oxford)*. 2010, 49: 2327–36.
- Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ*. 2009, 339: b2535.
- 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*. 2010a, 19: 539–49.
- 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*. 2010b, 63: 737–45.
- Nielke MC, Lindenhovius AL, Watson JB, Vranceanu AM, Ring D. Correlation of DASH and QuickDASH with measures of psychological distress. *J Hand Surg Am*. 2009, 34: 1499–505.
- Rogers MW, Wilder FV. The effects of strength training among persons with hand osteoarthritis: A two-year follow-up study. *J Hand Ther*. 2007, 20: 244–9.
- Rogers MW, Wilder FV. Exercise and hand osteoarthritis symptomatology: a controlled crossover trial. *J Hand Ther*. 2009, 22: 10–7.
- Rosales RS, Diez de la Lastra I, McCabe S, Ortega Martinez JI, Hidalgo YM. The relative responsiveness and construct validity of the Spanish version of the DASH instrument for outcomes assessment in open carpal tunnel release. *J Hand Surg Eur*. 2009, 34: 72–5.
- Rudolf KD, Kus S, Coenen M, Dereskewitz C, van de Ven-Stevens LAW, Cieza A. Report on the International ICF Consensus Conference on the ICF core sets for hand conditions. *Hand Therapy*. 2010, 15: 73–6.
- Rudolf KD, Kus S, Chung KC, Johnston M, LeBlanc M, Cieza A. Development of the International Classification of Functioning, Disability and Health core sets for hand conditions—results of the World Health Organization International Consensus process. *Disabil Rehabil*. 2012, 34: 681–93.
- Sanderson S, Tatt ID, Higgins JP. Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: a systematic review and annotated bibliography. *Int J Epidemiol*. 2007, 36: 666–76.
- Schulz KF, Altman DG, Moher D. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *BMC Med*. 2010, 8: 18.
- Scientific Advisory Committee of the Medical Outcomes Trust. Assessing health status and quality-of-life instruments: attributes and review criteria. *Qual Life Res*. 2002, 11: 193–205.
- Spaans AJ, van Laarhoven CMCM, Schuurman AH, van Minnen LP. Interobserver agreement of the Eaton-Littler classification system and treatment strategy of thumb carpometacarpal joint osteoarthritis. *J Hand Surg Am*. 2011, 36A: 1467–70.
- Stamm T, van der Giesen F, Thorstensson C et al. Patient perspective of hand osteoarthritis in relation to concepts covered by instruments measuring functioning: a

- qualitative European multicentre study. *Ann Rheum Dis*. 2009, 68: 1453–60.
- Terwee CB, Bot SD, de Boer MR et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol*. 2007, 60: 34–42.
- Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res*. 2012, 21: 651–7.
- Uchiyama S, Imaeda T, Toh S et al. Comparison of responsiveness of the Japanese Society for Surgery of the Hand version of the carpal tunnel syndrome instrument to surgical treatment with DASH, SF-36, and physical findings. *J Orthop Sci*. 2007, 12: 249–53.
- Valderas JM, Ferrer M, Mendivil J et al. Development of EMPRO: a tool for the standardized assessment of patient-reported outcome measures. *Value Health*. 2008, 11: 700–8.
- Valdes K, Marik T. A systematic review of conservative interventions for osteoarthritis of the hand. *J Hand Ther*. 2010, 23: 334–50.
- Vandenbroucke JP, von Elm E, Altman DG et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *Epidemiology*. 2007, 18: 805–35.
- Vermeulen GM, Slijper H, Feitz R, Hovius SE, Moojen TM, Selles RW. Surgical management of primary thumb carpometacarpal osteoarthritis: a systematic review. *J Hand Surg Am*. 2011, 36: 157–69.
- Wajon A, Carr E, Edmunds I, Ada L. Surgery for thumb (trapeziometacarpal joint) osteoarthritis. *Cochrane Database Syst Rev*. 2009: CD004631.
- Wilder FV, Barrett JP, Farina EJ. Joint-specific prevalence of osteoarthritis of the hand. *Osteoarthritis Cartilage*. 2006, 14: 953–7.
- Wright JG, Swiontkowski MF, Heckman JD. Introducing levels of evidence to the journal. *J Bone Joint Surg Am*. 2003, 85-A: 1–3.