Journal of Hand Therapy xxx (xxxx) xxx

Contents lists available at ScienceDirect

Journal of Hand Therapy

journal homepage: www.elsevier.com/locate/jht

Patient-reported outcome measures used for hand and wrist disorders: An overview of systematic reviews

Christina Ziebart, PT, PhD[a,b,c,*], Pavlos Bobos, PT, PhD[a,b,c,e], Rochelle Furtado, PT, MSc, PhD (c)[a,b,c], Armaghan Dabbagh, PT, MSc, PhD (c)[a,c], Joy MacDermid, PT, PhD[a,b,c,d,f]

[a] Department of Rehabilitation Sciences, Faculty of Health Science, Western University, London, Ontario, Canada
[b] School of Physical Therapy, Faculty of Health Science, Western University, London, Ontario, Canada
[c] Collaborative Program in Musculoskeletal Health Research, Bone and Joint Institute, Western University, London, Ontario, Canada
[d] Roth McFarlane Hand and Upper Limb Centre, St. Joseph's Hospital, London, Ontario, Canada
[e] Dalla Lana School of Public Health, Department of Clinical Epidemiology and Health Care Research, Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, Ontario, Canada
[f] School of Rehabilitation Science, McMaster University, Hamilton, Ontario, Canada

ARTICLE INFO

Article history:
Received 31 March 2021
Revised 19 May 2022
Accepted 14 October 2022
Available online xxx

Keywords:
Patient reported outcomes
Hand
Wrist
Pain

ABSTRACT

Background: Multiple options for patient reported outcome measures are available to assess patients with hand, wrist and elbow impairments. This review of systematic reviews (overview) evaluated the evidence on these outcome measures.
Methods: An electronic search of six databases (MEDLINE, Embase, CINAHL, ILC, the Cochrane Central Register of Controlled Trials (CENTRAL), and LILACS) was performed in September 2019, and updated in August 2022. The search strategy was designed to locate systematic reviews that addressed at least one clinical measurement property of PROMs used for patients with hand and wrist impairment. Two independent reviewers screened the articles and extracted the data. The AMSTAR tool was used to assess the risk of bias in the included articles.
Results: Eleven systematic reviews were included in this overview. A total of 27 outcome assessments were assessed, with DASH, PRWE and MHQ assessed by five, four, and three reviews, respectively. We found high-quality evidence of good to excellent internal consistency (ICC = 0.88-0.97), poor content validity but high construct validity (r > 0.70), moderate- to high-quality evidence for the DASH. The reliability of the PRWE was excellent (ICC >0.80), the convergent validity was excellent (r > 0.75), but poor criterion validity compared to the SF-12. The MHQ also reported excellent reliability (ICC = 0.88-0.96), and good criterion validity (r > 0.70), but poor construct validity (r > 0.38).
Conclusion: Clinical decisions around which tool will depend on which psychometric property is most important for the assessment and whether global or specific condition assessment is needed. All of the tools demonstrated at least good reliability; therefore, the clinical decisions will rely on the type of validity for clinical application. The DASH has good construct validity, while the PRWE has good convergent validity, and the MHQ has good criterion validity.

Introduction

Patient-reported outcome measures (PROMs) are considered the favored assessment by clinicians in upper extremity orthopedic conditions.[1] They provide insight into a patient's experience of their health problem, regarding their upper extremity disorder. PROMs can identify problematic activities of daily living, pain, function and sensation, which can be incorporated into a more patient-centered treatment plan.[2,3]

Hand and wrist PROMs are frequently used in the management of hand and wrist conditions, leading to the creation of different PROMs evaluating various hand, wrist and elbow conditions.[4,5] For example, there are some PROMs that are general to the whole upper extremity, like the Disability of the Arm, Shoulder and Hand (DASH) score, and others that are more region specific to the hand and wrist, like the Patient Rated Wrist Evaluation (PRWE), or the

Conflict of interest: All named authors hereby declare that they have no conflicts of interest to disclose.

* Corresponding author. Faculty of Health Science, Western University, 1201 Western Rd, London, ON N6G 1H1, Canada.
E-mail address: cziebart@uwo.ca (C. Ziebart).

0894-1130/$ – see front matter © 2022 Elsevier Inc. All rights reserved.
https://doi.org/10.1016/j.jht.2022.10.007

elbow, like the Patient Rated Elbow Evaluation (PREE).[6] PROMs are often used by clinicians to determine a patient's status at the time of the assessment, to predict a subsequent event, or to monitor and detect change over time.[7] Therefore, the tool needs to be both reliable and valid.[8] Implementation of assessment tools in practice is low,[9] and this may be due to difficulty understanding which assessment tools' most appropriate.[10]

Recognizing the large volume of outcome assessments used for hand, wrist and elbow conditions can be challenging when understanding (1) the strengths and limitations of the measure and (2) selection of the appropriate measure for the patient. Individual studies evaluating the psychometric properties of instruments are often summarized into systematic reviews, which can provide a useful synthesis of a larger body of evidence on hand, wrist and elbow PROMs.[11,12] However, with the abundant number of systematic reviews, evaluating and reaching a single consensus on the utility of a PROM is lacking and challenging to navigate in the literature.

An overview which synthesizes systematic reviews, can provide a more fulsome synthesis of the literature in a single manuscript.[13] The primary purpose of this study was to conduct an overview of systematic reviews and synthesize evidence to establish the current state of knowledge on clinical measurement properties of PROMs for patients with hand, wrist and elbow conditions. The secondary purpose was to assess the primary studies for heterogeneity regarding their clinical measurement properties to better establish consensus of PROMs.

Methods

Study design

The study design is an overview of systematic reviews. An overview is a method of synthesizing the findings from multiple systematic reviews.[13] This study has been registered on PROSPERO. Registration: CRD 42019137491

Search method

An electronic search of six databases (MEDLINE, Embase, CINAHL, ILC, the Cochrane Central Register of Controlled Trials (CENTRAL), and LILACS) was performed in September 2019, and updated in August 2022 for relevant literature for inclusion from the respective inception dates of the database. No other restrictions were imposed, and grey literature was permitted. The search strategy was designed to locate systematic reviews that addressed at least one clinical measurement property of PROMs used for patients with hand, wrist, or elbow condition. The search strategy, including keywords and Boolean operators is shown in appendix 1.

Study selection

Retrieved articles were entered into EndNote X9 (Clarivate Analytics, Boston, MA), and reviewed independently by two authors (PB and CZ). Titles and abstracts were reviewed, and the following inclusion and exclusion criteria were applied:

Inclusion criteria

- assessed at least one outcome measure for hand, wrist, or elbow condition.
- at least one of the following clinical measurement properties: validity, reliability, responsiveness, Rasch analysis, factor analysis, cross-cultural validation, interpretability, and floor/ceiling effect.

Exclusion criteria

- Studies that did not mention any clinical measurement properties for the outcome measures.
- Non-English Text.
- No full text available.

Risk-of-bias assessment

Two review authors (CZ and PB) applied *A MeaSurement Tool to Assess systematic Reviews* (AMSTAR).[14] The 11-item AMSTAR tool is used for assessing the risk of bias in systematic reviews. Reviews that achieve a score of 8 or higher were considered low risk, reviews scoring between five and seven were moderate risk of bias, and studies scoring less than four were high risk of bias.[14] The first version of the AMSTAR was used because of it having better psychometric properties than the AMSTAR-2.[15]

Data extraction

The data-extraction form was adapted from a previous overview, assessing similar outcome measurements.[16] Two authors performed the data extraction (RF and AD), verified by the first author (CZ). Both data and descriptive elements of the study were extracted. Descriptively, the sample size, patient population, purpose of the study and clinical measurement properties investigated were extracted. Data related to the clinical measurement properties presented in the systematic reviews were extracted; data from the original articles within each systematic review was not extracted.

Data synthesis

A qualitative synthesis was conducted to report findings on clinical measurement properties. High quality evidence was defined as similar findings reported in at least two low-to-moderate risk of bias systematic reviews (SRs), moderate quality evidence was defined as one or more moderate risk of bias SRs with similar findings, with or without conflicting findings from high risk of bias reviews, and low-quality evidence was defined 1 or more high risk of bias reviews. Conflicting evidence was defined as when reviews of similar quality reported different conclusions.

Results

Initially, 55 articles were identified by the electronic database search. After removing the duplicates and the first phase of screening, ten reviews were evaluated in the full text. After the second phase of screening (full-text screening), six SRs were eligible for inclusion in this overview. Through a second data search and snowballing methods, an additional five articles were identified, resulting in a total of 11 articles included in this overview. The information on the psychometric properties of the PROMs presented by these SRs, were used for a narrative synthesis of the results of this overview. Fig. 1 presents the PRISMA diagram of the results of the database search and screening phases. In all these SRs, a sample of people with hand, wrist, or elbow injuries (both musculoskeletal and neurological) were included, and the psychometric properties of at least one PROM was reviewed. Table 1 summarizes the characteristics of the included SRs, outlining the number of included studies in each SR, the patient population, and the assessed PROM(s).

Overall, 27 different PROMs were reviewed by the included SRs (Table 2). These PROMs were as follows: Disabilities of the Arm, Shoulder and Hand (DASH), quick DASH (QDASH), the simplified Chinese version of DASH (DASH-CHNPLAGH), Patient-Rated Wrist and Hand Evaluation/ Patient Rate Wrist Evaluation (PRWHE/ PRWE), Michigan Hand Outcome Questionnaire (MHQ), Patient

Please cite this article as: C. Ziebart, P. Bobos, R. Furtado et al., Patient-reported outcome measures used for hand and wrist disorders: An overview of systematic reviews, Journal of Hand Therapy, https://doi.org/10.1016/j.jht.2022.10.007

Table 1
Description of the studies.

Author & year	Number of studies	PROM(s) evaluated	Study population
Bialocerkowski et al. 2000	9	PRWE, Pain VAS	MSK wrist injuries
Changulani et al., 2008	6	DASH, Brigham and Women's Hospital CTQ, PRWE, Gartland and Werley score	Any population
Dabbagh et al. 2020	12	Bland questionnaire, Kamath and Stothard questionnaire, CTS-6, BCTQ	CTS
de Klerk et al. 2018	14	DASH	MSK hand injuries
Fonseca et al. 2018	16	CISS, CSS, PWES, PEM, MHQ, DASH, HAT, SF-12, PRWHE, DASH-CHNPLAGH	Median nerve injury, ulnar nerve injury, nerve injury/repair, radial nerve injury, digital nerve injury
de Carvalho Leite et al. 2006	10	BCTQ	CTS
McPhail, 2012	13	DASH, Pain VAS, ADL Questionnaire, Hand function VAS, Thumb pain VAS, MHQ, SF12	Wrist OA
Mehta et al. 2015	22	PRWE	General hand/wrist conditions, ORIF, DRF, RIAP
Taylor, 2014	5	PRWE	MSK wrist injuries
Vincent et al. 2019	9	PREE, pASES-e	Total elbow arthroplasty, various elbow conditions, OA, RA of the elbow
Wormald et al. 2019	54	DASH, QDASH, MHQ, UEFI, PEM, DHI	CTS, hand surgery/trauma, scleroderma, dupuytren's

CISS = Cold Intolerance Symptom Severity; CSS = Cold Sensitivity Severity; PWES = Potential Work Exposure Scale; PEM = Patient Evaluation Measure; MHQ = Michigan Hand Outcome Questionnaire; DASH = Disabilities of the Arm, Shoulder and Hand; DASH-CHNPLAGH = simple Chinese version of DASH; PRWHE = Patient-Rated Wrist and Hand Evaluation; HAT = hand assessment tool; SF12 = Health Survey (Short Form 12); DRF = distal radius fracture; RIAP = resection interposition arthroplasty; DHI = Duruoz hand index; UEFI = Upper Extremity Functional Index; PREE = Patient-Rated Elbow Evaluation; pASES-e = patient-reported section of the American Shoulder and Elbow Surgeons -elbow form

Table 2
Patient-reported outcome measures, categorized by the number of times the outcome was measured in either the systematic review or the primary study.

Patient-reported outcome measures	Number of systematic reviews (%)	Number of primary Studies (%)
DASH	5 (45)	130 (76)
PRWE	4 (36)	42 (25)
MHQ	3 (27)	83 (49)
PEM	2 (18)	70 (41)
SF-12	2 (18)	29 (17)
Pain VAS	2 (18)	22 (13)
BCTQ	2 (18)	22 (13)
ADL	1 (9)	13 (8)
Bland Questionnaire	1 (9)	12 (7)
CISS	1 (9)	16 (9)
CTS-6	1 (9)	12 (7)
CSS	1 (9)	16 (9)
DHI	1 (9)	54 (32)
DASH-CHNPLAGH	1 (9)	16 (9)
Gartland and Werley	1 (9)	6 (4)
Hand Function	1 (9)	13 (8)
Kamath and Stothard questionnaire	1 (9)	12 (7)
Thumb Pain	1 (9)	13 (8)
HAT	1 (9)	16 (9)
pASES-e	1 (9)	9 (5)
CTQ	1 (9)	6 (4)
PREE	1 (9)	9 (5)
PRWHE	1 (9)	16 (9)
PWES	1 (9)	16 (9)
QDASH	1 (9)	54 (32)
UEFI	1 (9)	54 (32)

DASH = Disabilities of the Arm, Shoulder and Hand; QDASH = quick DASH; DASH-CHNPLAGH, the simplified Chinese version of DASH; PRWHE = Patient-Rated Wrist and Hand Evaluation; MHQ = Michigan Hand Outcome questionnaire; PEM = Patient Evaluation Measure; BCTQ = Boston Carpal Tunnel questionnaire; CISS = Cold Intolerance Symptom Severity; CSS = Cold Sensitivity Severity; DHI = Duruoz Hand Index; HAT = Hand Assessment Tool; PRWHE = Patient-Rated Wrist and Hand Evaluation; PREE = Patient-Rated Elbow Evaluation; pASES-e = patient-reported section of the American Shoulder and Elbow Surgeons -elbow form; PWES = Potential Work Exposure Scale; SF-12 = Health Survey (Short Form 12); UEFI = Upper Extremity Functional Index.

Please cite this article as: C. Ziebart, P. Bobos, R. Furtado et al., Patient-reported outcome measures used for hand and wrist disorders: An overview of systematic reviews, Journal of Hand Therapy, https://doi.org/10.1016/j.jht.2022.10.007

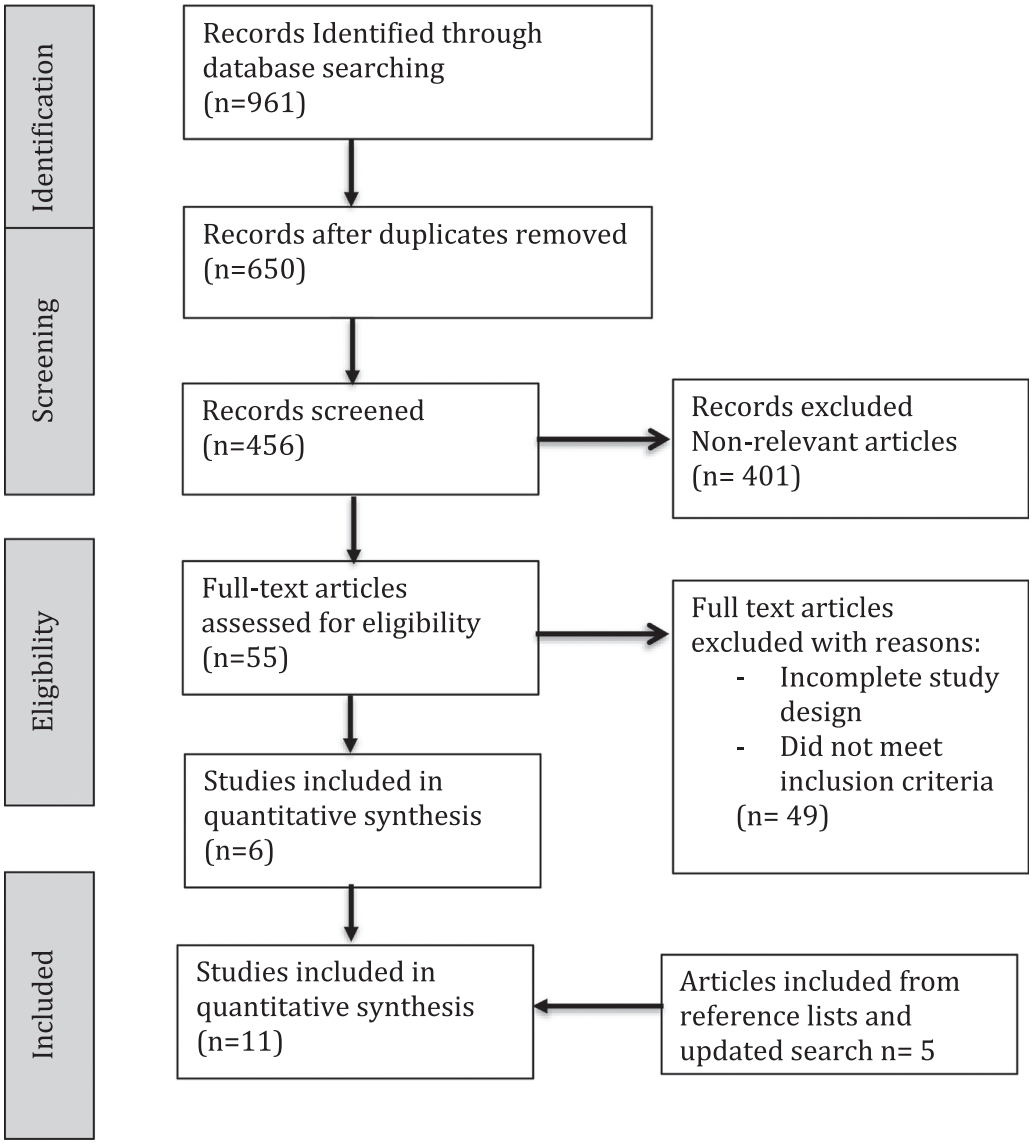

Fig. 1. Flow diagram of the selection of studies.

Evaluation Measure (PEM), Activities of Daily Living (ADL) questionnaire, Bland questionnaire, Boston Carpal Tunnel questionnaire (BCPT), Cold Intolerance Symptom Severity (CISS), Cold Sensitivity Severity (CSS), Carpal Tunnel Syndrome (CTS-6) Duruoz Hand Index (DHI), Hand Assessment Tool (HAT), Gartland and Werley, Hand function questionnaire, Kamath and Stothard questionnaire, Thumb Pain, Patient-Rated Elbow Evaluation (PREE), patient-reported section of the American Shoulder and Elbow Surgeons -elbow form (pASES-e), Potential Work Exposure Scale (PWES), Health Survey (Short Form 12) (SF12), and lastly, Upper Extremity Functional Index (UEFI), Pain Visual Analog Scale (VAS). The following sections elaborate on the risk of bias (ROB) of these SRs and an overview of the reported psychometric properties of the reported PROMs. The DASH, PRWE, and MHQ are discussed in their own paragraphs because they were highly cited PROMs.

Risk-of-bias assessment of the included systematic reviews

The assessment of the included SRs is summarized in Table 3. Overall, two of the included SRs had a low ROB,[17,18] and the remaining nine SRs had moderate ROB.[5,12,19-25] On the first item of the AMSTAR, asking whether a priori design was provided, two SR was rated as 'yes.' The second question on AMSTAR asked whether there was duplicate study selection and extraction and seven of the studies did, whereas four did not. On the items three and nine, all the SRs were rated as 'yes', meaning that all the SRs had performed a comprehensive database search, and the methods used in combining the results were appropriate. However, for question eight asking whether appropriate conclusions were made, one study did not. On the contrary to these items, we rated all the SRs as 'no' on items 10, which asked about the likelihood of publication bias. Question 11 asked about conflict of interested and only one study reported on the conflict of interest. On AMSTAR item four, two SRs had included the status of the publication (ie, grey literature), therefore we rated them as 'yes.' On item five, asking about a list of included and excluded studies, four of the SRs rated as 'yes' and the remaining as 'no.' Lastly, nine SRs had assessed the scientific quality of the included SRs and we rated them as 'yes' on the AMSTAR item seven.

Properties of specific patient-reported outcome measures

The Disability of the Arm, Shoulder and Hand (DASH)

The psychometric properties of DASH were reviewed by five SRs (130 primary studies), one SR with low ROB,[18] and four SRs with

Table 3
AMSTAR ratings.

Reviews	1	2	3	4	5	6	7	8	9	10	11	ROB rating
Bialocerkowski, 2000	N	N	Y	N	N	N	Y	Y	Y	N	N	Moderate
Changulani, 2008	N	N	Y	N	N	N	N	N	Y	N	N	Moderate
Dabbagh, 2020	Y	Y	Y	N	Y	Y	Y	Y	Y	N	Y	Low
de Klerk, 2018	N	Y	Y	Y	N	Y	Y	Y	Y	N	N	Moderate
Fonseca, 2018	N	Y	Y	N	Y	Y	Y	Y	Y	N	N	Moderate
de Carvalho Leite, 2006	N	Y	Y	N	N	Y	N	Y	Y	N	N	Moderate
McPhail, 2012	N	N	Y	N	N	Y	Y	Y	Y	N	N	Moderate
Mehta, 2015	N	Y	Y	N	Y	Y	Y	Y	Y	N	N	Moderate
Taylor, 2014	N	N	Y	N	N	N	Y	Y	Y	N	N	Moderate
Vincent, 2019	N	Y	Y	N	Y	Y	Y	Y	Y	N	N	Moderate
Wormald, 2019	Y	Y	Y	Y	N	Y	Y	Y	Y	N	N	Low

1. Was an 'a priori' design provided?
2. Was there duplicate study selection and data extraction?
3. Was a comprehensive literature search performed?
4. Was the status of publication (ie, grey literature) used as an inclusion criterion?
5. Was a list of studies (included and excluded) provided?
6. Were the characteristics of the included studies provided?
7. Was the scientific quality of the included studies assessed and documented?
8. Was the scientific quality of the included studies used appropriately in formulating conclusions?
9. Were the methods used to combine the findings of studies appropriate?
10. Was the likelihood of publication bias assessed?
11. Was the conflict of interest included?

Table 4
Summary of the psychometric properties of DASH.

Instrument/ Review	Reliability	Validity	Responsiveness	ROB
DASH/ Wormald et al. 2019	Internal consistency in CTS = 0.97, trauma = 0.98, RA = 0.98; Test-retest reliability in trauma ICC = 0.98, CTS ICC = 0.87, RA: ICC = 0.97	Poor content validity: For the criterion validity, moderate correlation with the SF - 36 for surgical patients was reported; For construct validity, results were in accordance with hypothesis for trauma patients; For structural validity, failed Rasch analysis in both dupuytren's population and hand injuries	Responsiveness was good in patients of CTS and Trauma/ surgical, specific numbers not reported	Low
DASH/ Fonseca et al. 2018	Internal consistency = 0.98;	The PEM, MHQ, and DASH PROMS had a Pearson's correlation coefficient of r> 0.38 in nerve injury patients;	Not reported	Moderate
DASH/ De Klerk et al. 2018	Not reported	Content validity rarely reported amongst studies (8/14 studies); Good construct validity; Cross-cultural validity was excellent among all populations; Factorial validity: principal component analysis was done, justifying 7 of the subscales	Not reported	Moderate
DASH/ Changulani et al., 2008	Test-retest reliability correlation = 0.96	Construct validity determined through correlation with Brigham CTQ (0.73); SPADI (0.72) and pain severity (0.67)	SRM = 0.74	Moderate
DASH/ McPhail et al., 2012	Not reported	Not Reported	Not Reported	Moderate
QDASH/ Wormald et al. 2019	Internal consistency in trauma 0.92-0.95; test-retest reliability in trauma ICC = 0.90- 0.97	Poor content validity; For the criterion validity, good correlation (r>0.70) with DASH in trauma patients; For construct validity, result was in accordance with the hypothesis in dupuytren's patients;	Not reported	Low
DASH-CHNPLAGH/ Fonseca et al. 2018	Internal consistency: 0.88-0.96;	Negative correlation with SF-36, and positive correlation with VAS;	Not reported	Moderate

DASH = disability of the arm, shoulder, and hand; QDASH = quick DASH; DASH-CHNPLAGH, simple Chinese version of DASH; VAS = visual analogue scale; CTS = carpal tunnel syndrome; RA = rheumatoid arthritis; ICC = intraclass correlation coefficient; SF-36 = health survey short form-36; PEM = Patient Evaluation Measure; MHQ = Michigan Hand Outcome questionnaire

moderate ROBs.[5,20-22] Additionally, two other versions of the DASH were reviewed in the included SRs: Quick DASH (QDASH)[18] and a simple Chinese version of DASH (DASH-CHNPLAGH).[22]

Reliability of the DASH was reviewed in three of the included SRs.[18,20,22] High quality evidence supports excellent internal consistency and test-retest reliability of the DASH in carpal tunnel syndrome (CTS), hand rheumatoid arthritis (RA), and traumatic/surgical patients (Table 4).[18] Wormald et al. in an SR with low ROB, reported internal consistencies of 0.97, 0.98, and 0.98 for patients with CTS, trauma, and RA, respectively.[18] These results were similar to the findings of another SR with moderate ROB, that reported an overall internal consistency of 0.98.[22] Additionally, Wormald et al. reported the test-retest reliability of the DASH in terms of intraclass correlation coefficients (ICCs).[18] These ICCs were 0.87, 0.98, and 0.97 for CTS, trauma, and hand RA, respectively.[18] Wormald et al. also reported the internal consistency and test-retest reliability of the Q-DASH in hand trauma patients. In this SR internal consistency was excellent, with Cronbach's alpha

Please cite this article as: C. Ziebart, P. Bobos, R. Furtado et al., Patient-reported outcome measures used for hand and wrist disorders: An overview of systematic reviews, Journal of Hand Therapy, https://doi.org/10.1016/j.jht.2022.10.007

Table 5
Summary of the psychometric properties of PRWE.

Instrument/ Review	Reliability	Validity	Responsiveness	ROB
PRWE/ Bialocerkowski et al., 2000	Not Reported	Not Reported	Not Reported	Moderate
PRWE/ Changulani et al., 2008	Test-retest reliability Excellent ICC >0.90 in DRF population Functional subscore ICC = 0.85 Moderate reliability in scaphoid fracture ICC = 0.61	Construct Validity showing an improvement of 74% compared to the SF-36 which showed an improvement of 14% Criterion Validity correlation between SF-36 between 0.33 and 0.73; and weak correlation (-0.52) with a PROM score	Not Reported	Moderate
PRWE/ Mehta et al., 2015	ICC ranged from 0.81 to 0.94	Content validity item completeness was 78%-84% Construct validity r = >0.70 with DASH r = 0.3-0.7 with wrist PROM score r < 0.3 with wrist ROM Internal consistency greater than 0.75	SEM ranged from 5.2 to 8.1	Moderate
PRWE/ Taylor et al., 2014	Test re-test reliability high correlation ICC = 0.92	Good face validity High convergent validity r> 0.75	SRM: 1.51 ES: 1.61	Moderate

values ranging from 0.92 to 0.95 and an excellent test-retest reliability (ICC ranging from 0.90 to 0.97).[18] Fonseca et al., in a SR with moderate ROB, reported an excellent internal consistency ranging from 0.88 to 0.96 for the DASH-CHNPLAGH.[22]

The validity of the DASH was reviewed by three SRs, one with low ROB[18] and two with moderate ROBs.[21,22] High quality evidence reported poor (or rarely reported) content validity, good construct validity, poor Rasch analyses, and moderate convergent construct validity (Table 4).[18] Moderate quality evidence reported excellent cross-cultural validity.[21] A Pearson correlation coefficient of r>0.38 with PEM and the MHQ, reports poor validity.[22] High quality evidence reports poor content validity, good criterion validity (r>0.70) in hand trauma patients, and good construct validity for the Q-DASH.[18] Negative correlation with SF-36 and positive correlation with the pain visual analogue scale (VAS) was reported by a moderate quality SR for DASH-CHNPLAGH.[22]

Lastly, the responsiveness of the DASH was evaluated in two SR (with low and moderate ROB). These SRs reported good responsiveness of the DASH in CTS and hand trauma patients.[18,20] For a summary of the psychometric properties of the DASH, QDASH, and DASH-CHNPLAGH PROMs, please refer to Table 4.

Overall, the systematic reviews consistently report good to excellent internal consistency, moderate to excellent test-retest reliability, in studies with moderate- to high-quality evidence for the DASH.[18,20,22]

Patient Rated Wrist Evaluation (PRWE)

The PRWE was evaluated in four of the SRs, with 42 primary studies (Table 5). The test re-test reliability was assessed in three of the SR[12,20,24] and was reported to be moderate in two SR[12,20] and high in the other SR.[24] Validity was reported in three of the four SR.[12,20,24] Construct and face validity were both reported to be good.[20] There was high convergent validity, reported as r> 0.75[24]. Responsiveness was reported in one SR, and the SRM was 1.51 and the effect size was 1.61[24]. All of studies that evaluated the PRWE were of moderate quality.

Overall, the systematic reviews consistently reported good to excellent reliability, and weak to moderate validity, in studies with moderate quality of evidence for the PRWE.[12,20,24]

Michigan Hand Questionnaire (MHQ)

Another commonly reviewed hand PROM was the MHQ. MHQ was assessed in three SRs (83 primary studies) with low and moderate ROBs.[5,18,22] High quality evidence supports excellent internal consistency of the MHQ in trauma and hand RA patients, with a Cronbach's alpha of 0.93 and 0.88, respectively.[18] Moderate quality evidence supports excellent internal consistency of the MHQ in patient with median, ulnar, radial, or digital nerves injuries/repairs (Table 6).[22] Moreover, high quality evidence suggests good content validity, good criterion validity (r>70 correlation with DASH) in trauma patients, and good convergent validity of the MHQ.[18] Regarding the construct validity of the MHQ, Fonseca et al. reported that the MHQ, PEM, and DASH PROMs had a Pearson's correlation coefficient of r>0.38.[22] The responsiveness of the MHQ was described as good in CTS and traumatic/surgical patients by Wormald et al. [18]More information on the psychometric properties of the MHQ can be found in Table 6.

Overall, we found moderate to high quality of evidence of high reliability. The validity of the MHQ was poor to moderate, but the responsiveness was good[5,18,22].

Other hand PROMs

The following PROMs were reviewed by two SRs: BCTQ, PEM, SF-12 and Pain VAS. The BCTQ reported high internal consistency (0.80-0.90), and excellent test re-test reliability (r = 0.91-0.93). The BCTQ has moderate face and content validity (r = 0.87-0.90, compared to DASH), but a low correlation between the CTQ and clinical sensory tests (Table 7).[17,23] For the PEM, only the internal consistency was reviewed in the included SRs.[18,22] High quality evidence supports an excellent internal consistency of 0.94 in CTS and trauma patients.[18] Moderate quality evidence supports excellent internal consistency of the PEM in patient with median, ulnar, radial, or digital nerves injuries/repairs, with values ranging from 0.88 to 0.96.[22] The validity of the PEM as assessed by high quality evidence, suggests low content validity (developed based on experts' opinion only), good criterion validity (correlation was good with clinical assessments of ROM, grip strength, tenderness, swelling, r<0.70), and moderate construct validity (Table 7).[18] The SF-12 had good internal consistency: 0.88-0.96 in patient with median, ulnar, radial, or digital nerves injuries/repairs.

Psychometric properties of several other elbow, wrist, and hand PROMs were reviewed once within a SR. These papers addressed the following instruments (see Table 2 for number of primary studies). ADL questionnaire, Hand Function VAS, Thumb pain VAS, Bringham and Women's Hospital CQ, Bland Questionnaire, CISS, CTS-6, CSS, DHI, Garland and Werley score, HAT, pASES-e, PREE, PRWHE, PWES, SF12, and UEFI can be found in Table 7.[12,18,22,23,25]

Please cite this article as: C. Ziebart, P. Bobos, R. Furtado et al., Patient-reported outcome measures used for hand and wrist disorders: An overview of systematic reviews, Journal of Hand Therapy, https://doi.org/10.1016/j.jht.2022.10.007

Table 6
Summary of the psychometric properties of the Michigan Hand Questionnaire (MHQ) in the included SRs.

Instrument/ Review	Reliability	Validity	Responsiveness	ROB
MHQ/ Wormald et al. 2019	Internal consistency in trauma patients = 0.93, in RA patients = 0.88	Content validity: good performance of using previous literature to create questionnaire; Criterion validity: correlation was good with the DASH (r > 0.70) in trauma patients; Convergent validity: correlations with the Levine symptom score, the DASH (r > o.7o) for trauma/surgical patients	Responsiveness was good in patients with CTS and trauma/ surgical specific numbers not reported	Low
MHQ/ Fonseca et al. 2018	Internal consistency: 0.88-0.96 in patient with median, ulnar, radial, or digital nerves injuries/repairs	Construct validity: The PEM, MHQ, and DASH PROMs had a Pearson's correlation coefficient of > 0.38	Not reported	Moderate
MHQ/ McPhail et al., 2012	Not Reported	Not Reported	Not Reported	Moderate

ROM = range of motion; MHQ = Michigan Hand Outcome questionnaire; PEM = Patient Evaluation Measure; PROM = patient-reported outcome measure; CTS = carpal tunnel syndrome; DASH = disability of the arm, shoulder, and hand; RA = rheumatoid arthritis

Discussion

This overview synthesized evidence to establish the current state of knowledge on clinical measurement properties of PROMs for patients with hand, wrist, and elbow conditions. A total of 11 systematic reviews were summarized in this overview, reporting on 27 different PROMs.[12,18,21-23,25,26] This overview was able to identify tools with strong and weak psychometric properties in a variety of hand, wrist, and elbow patient populations, allowing for a reference document for clinicians to guide which tool is most appropriate for assessing a condition. The quality of the studies was moderate to high. The moderate to high quality was due to most studies including duplicate study selection and extraction, a comprehensive literature search, scientific quality of the included studies were assessed, appropriate conclusions were drawn from the scientific quality of the studies and the methods to combine the results were appropriate. Reduced quality was mostly due to reporting issues, such as not registering the review, making it difficult to determine whether decisions were made a priori. Studies were also downgraded due to not reporting conflicts of interest and not including a list of excluded studies. These limitations should be considered when assessing the quality of the studies and the reliability of the results. Overall, the studies were moderate to high quality, suggesting that the evidence reported is likely true and can be used to inform clinical and research decision-making around selection of tools for a population.

The quality of the systematic reviews included in this overview were assessed using the AMSTAR quality assessment tool.[14] According to the AMSTAR tool, most of the systematic reviews were rated as moderate, with only two systematic reviews rated as low risk of bias.[17,18] Many of the systematic reviews were down rated for not registering their review, thereby making it challenging to assess the a priori plan for the review. Study registration is a relatively straightforward process that allows for transparency and maintaining academic integrity. Further, only one of the studies reported a conflict of interest or a list of excluded studies, which is necessary for transparency, replication, and maintenance of academic integrity.

The Disability of the Arm, Shoulder and Hand (DASH) was the most reported outcome measure assessed in the systematic reviews.[18,21,22] The reviews found good reliability, and poor validity. One primary study in the systematic review evaluated the responsiveness, which was good. These psychometric properties are reassuring, because the DASH is a global measure of upper extremity disorders, meaning that it could be implemented clinically or academically across a variety of participants, while still having good psychometric properties.[18,20,21] Choosing a global outcome measure may be beneficial in a clinic to reduce patient and administrative burden from having to administer many questionnaires.

The PRWE and MHQ were also commonly reported tools, assessed in four and three systematic reviews, respectively. The PRWE showed excellent reliability and moderate to excellent validity. The MHQ reported good validity in one study, but poor validity in another study. It may be that the validity is more robust in more condition-specific questionnaires such as the PRWE and MHQ, as the PEM showed poor validity which is more of a global assessment tool. A decision around which questionnaire to use will require careful thought on consideration of the condition and the psychometric property. A tool with good validity and reliability would be best for clinical practice, which was seen in the PRWE and DASH. Structural validity and responsiveness would better indicate clinically meaningful change of a condition, which is likely more relevant to a clinician and to the patients. Responsiveness was poorly assessed in this clinical population and would be an important direction for future research on psychometric properties of patient reported outcome measures.

There were several conditions and injury types represented by this overview. Most commonly, CTS and nerve injuries were discussed. Most of the original systematic reviews included a broader population of musculoskeletal hand injuries, hand surgery or general hand and wrist conditions. One study included an elbow arthroplasty population.[25] Specific population groups seemed to be lacking, making it potentially unclear which tool would be best for a certain population, based on the psychometric properties. For example, this overview did not provide a clear understanding of which PROM would be best for distal radius fractures. The DASH, MHQ and PRWE have been reportedly used in a distal radius fracture population,[9,27] but it seems objective outcomes like range of motion, grip strength and pinch strength are preferred.[28] As well, scaphoid fractures were not addressed in the SRs, but previous literature has used DASH, PEM and VAS pain to assess PROM in scaphoid fractures.[29] Clinically, when deciding about which outcome measure to use, several factors should be considered.

The population of interest should be considered. Assessing the population will help to determine if there is a population-specific outcome measure, or whether a global upper extremity measure should be used. For example, someone with carpal tunnel syndrome may be better assessed with a carpal tunnel specific

Please cite this article as: C. Ziebart, P. Bobos, R. Furtado et al., Patient-reported outcome measures used for hand and wrist disorders: An overview of systematic reviews, Journal of Hand Therapy, https://doi.org/10.1016/j.jht.2022.10.007

Table 7
Summary of the psychometric properties of other hand, wrist and elbow PROMs.

Instrument/ Review	Reliability	Validity	Responsiveness	ROB
BCTQ/ Leite et al. 2006	Internal consistency: alpha = 0.80-0.90 for the SSS component and 0.88-0.93 for the FSS component; Test-retest reliability Pearson's r = 0.91-0.93	Good face and content validity; Construct validity: BCTQ vs DASH: r = 0.87- 0.90 (strong), BCTQ vs AIMS-2: r = 0.70 (moderate), BCTQ vs generic health measures: r = 0.50- 0.56 (moderate), weak correlation between BCTQ and clinical sensory tests (r = 0.15- 0.17 for SSS, and r = 0.24- 0.42 for FSS), moderate to high correlation of pinch and grip strengths with FSS component, and strong correlation of sensibility measures with the SSS component	Both components yielded moderate (>0.5) to large (>0.8) responsiveness indices, thus both scales are sensitive to clinical change in patients undergoing surgical interventions	Moderate
BCTQ/ Dabbagh et al., 2020	Not Reported	Sensitivity of 35.1% for the FSS and 48.6 for the SSS components. Specificity of 62.5% for the FSS and 60% for the SSS components	Not Reported	
Bland Questionnaire/ Dabbagh et al., 2020	Not Reported	Sensitivity ranged from 78% for web-based to 82% for the paper-based versions. Specificity ranged from 55.6% for the web-based to 67% for the paper-based versions.	Not Reported	Low
Bringham and Women's Hospital Questionnaire/ Changulani et al., 2008	Test re-test reliability correlation 0.91 for symptom severity scale and 0.93 for functional status Internal consistency was 0.89 for symptom severity and 0.91 for functional scale	Moderate correlation with grip and pinch strength	ES: 1.4 for clinical change ES: 0.82 for functional score	Moderate
CISS/ Fonseca et al. 2018	Internal consistency: 0.88-0.96 in patient with median, ulnar, radial, or digital nerves injuries/repairs; Test-retest reliability: ICC = 0.85	Very strong content validity: The cold sensitivity PROM had a Spearman's correlation of 0.73 for CISS and 0.67 for CSS for construct validity	Not reported	Moderate
CTS-6/ Dabbagh, et al., 2020	Not Reported	Sensitivity ranged from 87% to 95% and specificity ranged from 60% to 91%	Not Reported	Low
CSS/ Fonseca et al. 2018	Internal consistency: 0.88-0.96 in patient with median, ulnar, radial, or digital nerves injuries/repairs; Test-retest reliability: ICC = 0.85	Very strong content validity: The cold sensitivity PROM had a Spearman's correlation of 0.73 for CISS and 0.67 for CSS for construct validity;	Not reported	Moderate
DHI/ Wormald et al. 2019	Internal consistency: 0.7-0.98 in hand/ surgical patients; Test-retest reliability: ICC = 0.8-0.97 in scleroderma, ICC = 0.99 in other hand/surgical patients	Criterion validity: no correlation reported in scleroderma or surgical patients; Other types of validity were not reported in the literature.	potential responsiveness properties were measured in accordance with the hypothesis in hand/surgical patients	Low
HAT/ Fonseca et al. 2018	Internal consistency: 0.88-0.96 in patient with median, ulnar, radial, or digital nerves injuries/repairs	Not reported	Not reported	Moderate
Kamath and Stohard questionnaire / Dabbagh et al., 2020	Not Reported	Sensitivity ranged from 74% to 87% and specificity ranged from 64% to 87%	Not Reported	Low
Gartland and Werley questionnaire/ Changulani et al., 2008	Not Reported	Not Reported	Not Reported	Moderate
Pain VAS/ Bialocerkowski et al., 2000	Not Reported	Not Reported	Not Reported	Moderate

(*continued on next page*)

Table 7 (*continued*)

Instrument/ Review	Reliability	Validity	Responsiveness	ROB
pASES-e/ Vincent et al. 2019	Internal consistency: 0.93 in various elbow conditions, 0.9 in total elbow arthroplasty; Test-retest reliability ICC in total elbow arthroplasty: 0.73-0.94, in OA and RA: 0.64-0.90; MDC95 = 16.05 points, SEM = 1.62, 6.2, 11.62 (for each of the three subscales)	Construct validity: p-ASES-e vs SF36 was r>0.70; Longitudinal validity: PREE vs pASES-e = 0.74 (p<0.01), PREE vs DASH = 0.62 (p0.01); p-ASES-e vs DASH= 0.23; p-ASES-e vs PREE= 0.33; Discriminant validity was established with the mental component of the SF-36; AUC = 0.67, sensitivity = 0.65, specificity = 0.69	ES= 1.55 in total elbow arthroplasty	Moderate
PWES/ Fonseca et al. 2018	Internal consistency: 0.88-0.96 in patient with median, ulnar, radial, or digital nerves injuries/repairs; Test-retest reliability: ICC= 0.85	Not reported	Not reported	Moderate
PREE/ Vincent et al. 2019	Internal consistency: 0.95 in various elbow conditions; Test-retest reliability ICC in total elbow arthroplasty: 0.90, in various elbow conditions: 0.73-0.94, in OA and RA: 0.93	Construct validity: PREE vs QDASH, DASH, pASES-e was r >0.70; Longitudinal validity: PREE vs pASES-e = 0.74 ($p <$ 0.01), PREE vs DASH = 0.62 (p0.01); p-ASES-e vs DASH= 0.23, p-ASES-e vs PREE= 0.33; Discriminant validity was established with the mental component of the SF-36; AUC = 0.68, sensitivity = 0.63, specificity = 0.71	ES = 1.7 and SRM = 1.37 in total elbow arthroplasty, ES = 1.32 and SRM = 1.28 in various elbow conditions; ceiling effect was observed	Moderate
PRWHE/ Fonseca et al. 2018	Internal consistency: 0.88-0.96 in patient with median, ulnar, radial, or digital nerves injuries/repairs	PRWHE was analyzed by Rasch analysis, supporting the internal consistency of the scale ($\alpha = 0.96$) and reliability (as measured by the person separation index) of 0.95	Not reported	Moderate
SF-12/ Fonseca et al. 2018	Internal consistency: 0.88-0.96 in patient with median, ulnar, radial, or digital nerves injuries/repairs	Not reported	Not reported	Moderate
UEFI/ Wormald et al. 2019	Test-retest reliability ICC= 0.85- 0.94	Content validity: expert opinion and literature review only, poor performance; Criterion validity: correlation was good with the DASH r > 0.70 in trauma patients; Construct validity: Result was accordance with the hypothesis with trauma patients; Structural validity: Rasch analysis did not support the validity of the 20-item UEFI so the UEFI15 was developed	Responsiveness was good in patients with CTS and trauma/ surgical, AUC = 0.88	Low
PEM/ Wormald et al. 2019	Internal consistency = 0.94 in CTS and trauma patients	Content validity: expert opinion only, poor performance; Criterion validity: correlation was good with clinical assessments of ROM, grip strength, tenderness, swelling r < 0.70; Construct validity: result was accordance with the hypothesis in trauma and CTS patients	Responsiveness was good in patients with CTS, and trauma/ surgical	Low
PEM/ Fonseca et al. 2018	Internal consistency: 0.88-0.96 in patient with median, ulnar, radial, or digital nerves injuries/repairs	Construct validity: The PEM, MHQ, and DASH had a Pearson's correlation coefficient of > 0.38	Not reported	Moderate

BCTQ = Boston Carpal Tunnel Questionnaire; CTS = carpal tunnel syndrome; AUC = area under the curve; ES = effect size; SRM, SSS = Symptom Severity Scale; FSS = Functional Status Scale; UEFI = upper extremity functional index; CISS = Cold Intolerance Symptom Severity; CSS = Cold Sensitivity Severity; PWES = Potential Work Exposure Scale; HAT = hand assessment tool, SF12 = Health Survey (Short Form 12); DHI = Duruoz hand index; PREE = Patient-Rated Elbow Evaluation; pASES-e = patient-reported section of the American Shoulder and Elbow Surgeons- elbow form; PRWHE = Patient-Rated Wrist and Hand Evaluation

Please cite this article as: C. Ziebart, P. Bobos, R. Furtado et al., Patient-reported outcome measures used for hand and wrist disorders: An overview of systematic reviews, Journal of Hand Therapy, https://doi.org/10.1016/j.jht.2022.10.007

questionnaire like the CTS-6, but someone with a distal radius fracture may be better assessed with the DASH, MHQ, or PRWE. Next, consider the purpose of the tool, whether it's to determine a patient's status at the time of the assessment, to predict a subsequent event, where good validity and reliability would be more important, or to monitor and detect change over time, where good responsiveness would be more important.[7,8] The DASH showed good reliability and responsiveness but poor validity, and the PREE showed good reliability and validity, but poor responsiveness (due to a ceiling effect).[25]

Generally, for a quick clinical decision, it seems that if someone presented with an upper extremity disorder the DASH might be a good first outcome measure to use, as it showed good validity in any population, as well as in more specific populations such as those with CTS, or after hand surgery or trauma. There are many outcome measures used for CTS, but the DASH, MHQ, and BCTQ demonstrate good responsiveness properties, indicating its usefulness in clinic. This overview contained both strengths and limitations that should be considered before the interpretation of our findings. This overview was able to synthesize a large amount of literature to guide clinicians and researchers on the appropriate choice of PROM for patients with a hand or wrist condition. However, it is important to note that while the systematic reviews had different objectives and eligibility criteria, the overlap in this synthesis and its conclusions were generally consistent amongst reviews. The PROMs that were included in this overview assessed at least one outcome measure for hand and wrist condition and at least one psychometric property. While there may be other PROMs not assessed in this overview, this overview does provide guidance on the mostly commonly used PROMs. Finally, while AMSTAR is a risk of bias tool to critically appraise the systemic review methods and not the adequacy of the measurement properties included in the reviews assessed.

Conclusion

Overall, the overview summarized the literature on PROMs for hand and wrist conditions. The DASH, PRWE and MHQ were the most reported PROM. The DASH had good reliability and responsiveness, poor content validity, but good construct validity, with moderate quality of evidence. The PRWE had excellent reliability and good convergent validity but poor construct validity, and the MHQ had good reliability, responsiveness, and good content validity but poor construct validity, with moderate to high quality of evidence. Clinical decisions may be made around the type of assessment required, whether it's for a global upper extremity disorder or a joint specific condition.

Acknowledgments

Christina Ziebart and Pavlos Bobos were supported by the CIHR doctoral award. Rochelle Furtado was supported by the Ontario Graduate Scholarship and the Bone and Joint Transdisciplinary Award. Joy C MacDermid was supported by a CIHR Chair in Gender, Work and Health and the Dr. James Roth Research Chair in Musculoskeletal Measurement and Knowledge Translation.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jht.2022.10.007.

References

1. Muenzen PM, Kasch MC, Greenberg S, et al. A new practice analysis of hand therapy. *J Hand Ther*. 2002;15:215–225.
2. Reflections on the practice of physiotherapy. *Scalpel (Brux)*. 1953;106:420–421.
3. Furtado R, Nazari G, MacDermid JC. A systematic review of the cross-cultural adaptations and measurement properties of the shoulder pain and disability index. *Hand Therapy*. 2019;24:107–115.
4. Dewan N, MacDermid JC, MacIntyre N. Validity and responsiveness of the short version of the Western Ontario rotator cuff index (short-WORC) in patients with rotator cuff repair. *J Orthopaed Sports Physic Ther*. 2018;48:409–418.
5. McPhail SM, Bagraith KS, Schippers M, et al. Use of condition-specific patient-reported outcome measures in clinical trials among patients with wrist osteoarthritis: a systematic review. *Advances orthoped*. 2012;2012.
6. MacDermid JC, Turgeon T, Richards RS, et al. Patient rating of wrist pain and disability: a reliable and valid measurement tool. *J Orthop Trauma*. 1998;12:577–586.
7. Guyatt GH, Feeny DH, Patrick DL. Measuring health-related quality of life. *Ann Intern Med*. 1993;118:622–629.
8. MacDermid JC, Stratford P. Applying evidence on outcome measures to hand therapy practice. *J Hand Ther*. 2004;17:165–173.
9. Michlovitz SL, LaStayo PC, Alzner S, et al. Distal radius fractures: therapy practice patterns. *J Hand Ther*. 2001;14:249–257.
10. MacDermid JC. Outcome evaluation in patients with elbow pathology: issues in instrument development and evaluation. *J Hand Ther*. 2001;14:105–114.
11. Mehta SP, MacDermid JC, Richardson J, et al. Reliability and validity of selected measures associated with increased fall risk in females over the age of 45 years with distal radius fracture - A pilot study. *J Hand Ther*. 2015;28:2–10.
12. Mehta SP, Macdermid JC, Richardson J, et al. A systematic review of the measurement properties of the patient-rated wrist evaluation. *J Orthop Sports Phys Ther*. 2015;45:289–298.
13. McKenzie JE, Brennan SE. Overviews of systematic reviews: great promise, greater challenge. *Syst Rev*. 2017;6 :Article number: 185.. doi:10.1186/s13643-017-0582-8.
14. Shea BJ, Bouter LM, Peterson J, et al. External validation of a measurement tool to assess systematic reviews (AMSTAR). *PLoS One*. 2007;2:e1350.
15. Lorenz RC, Matthias K, Pieper D, et al. AMSTAR 2 overall confidence rating: lacking discriminating capacity or requirement of high methodological quality? *J Clin Epidemiol*. 2020;119:142–144.
16. Bobos P, MacDermid JC, Walton DM, et al. Patient-Reported outcome measures used for neck disorders: an overview of systematic reviews. *J orthopaed sports physic ther*. 2018;48:775–788.
17. Dabbagh A, MacDermid JC, Yong J, et al. Diagnosing carpal tunnel syndrome: diagnostic test accuracy of scales, questionnaires, and hand symptom diagrams—a systematic review. *J orthopaed sports physic ther*. 2020;50:622–631.
18. Wormald JC, Geoghegan L, Sierakowski K, Price A, Peters M, Jain A, Rodrigues JN. Site-specific patient-reported outcome measures for hand conditions: systematic review of development and psychometric properties. *Plastic Reconstruct Surg Global Open*. 2019;7(5):e2256. PMID: 31333975; PMCID: PMC6571349. doi:10.1097/GOX.0000000000002256.
19. 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.
20. Changulani M, Okonkwo U, Keswani T, et al. Outcome evaluation measures for wrist and hand - Which one to choose? *Int Orthop*. 2008;32:1–6.
21. de Klerk S, Buchanan H, Jerosch-Herold C. The validity and clinical utility of the disabilities of the arm shoulder and hand questionnaire for hand injuries in developing country contexts: a systematic review. *J Hand Ther*. 2018;31:80–90 e81.
22. MdCR Fonseca, VMC Elui, Lalone E, et al. Functional, motor, and sensory assessment instruments upon nerve repair in adult hands: systematic review of psychometric properties. *Syst rev*. 2018;7:175.
23. de Carvalho Leite JC, Jerosch-Herold C, Song F. A systematic review of the psychometric properties of the boston carpal tunnel questionnaire. *BMC musculoskeletal disord*. 2006;7:78.
24. Taylor J, Kersten P. The patient-rated wrist and hand Evaluation: a systematic review of its validity and reliability. *New Zealand J Physiother*. 2014;42:139–145.
25. Vincent JI, MacDermid JC, King GJ, et al. Establishing the psychometric properties of 2 self-reported outcome measures of elbow pain and function: A systematic review. *J Hand Ther*. 2019;32:222–232.
26. Roy J-S, MacDermid JC, Woodhouse LJ. A systematic review of the psychometric properties of the Constant-Murley score. *J Shoulder Elbow Surg*. 2010;19:157–164.
27. Ikpeze TC, Smith HC, Lee DJ, et al. Distal radius fracture outcomes and rehabilitation. *Geriatr orthopaed surg rehab*. 2016;7:202–205.
28. Kumar S, Penematsa S, Sadri M, et al. Can radiological results be surrogate markers of functional outcome in distal radial extra-articular fractures? *Int Orthop*. 2008;32:505–509.
29. Lee C-H, Lee K-H, Lee B-G, et al. Clinical outcome of scaphoid malunion due to scaphoid fracture nonunion surgical treatment: a 5-year minimum follow-up study. *Orthopaed Traumatol: Surg Res*. 2015;101:359–363.

Please cite this article as: C. Ziebart, P. Bobos, R. Furtado et al., Patient-reported outcome measures used for hand and wrist disorders: An overview of systematic reviews, Journal of Hand Therapy, https://doi.org/10.1016/j.jht.2022.10.007