Hand Surgery and Rehabilitation 43 (2024) 101715

Available online at
ScienceDirect
www.sciencedirect.com

Elsevier Masson France
EM|consulte
www.em-consulte.com

Original article

Cross-cultural evaluation of the Michigan Hand Outcomes Questionnaire: a systematic review

María Visitación Martínez-Fernández [a], Irene Sandoval Hernandez [b], Jesús Martínez-Cal [c], Carmen Sarabia-Cobo [d,e,*]

[a] Rehabilitation Service, Hospital Mutua Montañesa, Avenida del Faro 19. Santander. 39012, Cantabria, Spain
[b] Department of Physical Therapy, Faculty of Health Science, University of Granada-Campus of Melilla, Calle Santander, 1, 52005 Melilla, Spain
[c] Department of Nursing, Physiotherapy and Medicine, Faculty of Health Sciences, University of Almería. 04120 Almería, Spain
[d] Department of Nursing, Universidad de Cantabria, Avenida de Valdecilla, s/n, 39008 Santander, Spain
[e] IDIVAL-Instituto de Investigación Sanitaria Valdecilla, s/n, Calle Cardenal Herrera Oria, 39011 Santander, Spain

ARTICLE INFO

Article history:
Received 29 December 2023
Received in revised form 6 April 2024
Accepted 8 April 2024
Available online 21 May 2024

Keywords:
Cross-cultural evaluation
Michigan Hand Outcomes Questionnaire
Systematic review

ABSTRACT

This systematic review aims to provide a comprehensive synthesis and in-depth analysis of the quality of the different cross-cultural versions of the MHQ. This study was conducted using Pubmed, Web of Science, CINAHL and SCOPUS databases to identify cross-cultural validation studies of the MHQ. Methodological quality, quality of evidence and criteria for good measurement properties of these studies were applied for each psychometric property. Quality assessment and data extraction were performed independently by two reviewers according to the COnsensus-based Standards for selection of health Measurement INstruments (COSMIN) guidelines. A total of 493 articles were identified, of which 22 were included and 20 were analysed.Of the six properties analysed, responsiveness and hypothesis testing for construct validity had the highest methodological quality and quality of evidence, and met the criteria for good measurement properties. The lowest quality properties were measurement error and internal consistency. The different cross-cultural versions of the MHQ were found to be reliable, valid and able to detect clinical change. The lack of development of measurement error, formulation of an a priori hypothesis or structural validity affects the detection of small clinical changes and their discriminative capacity.

1. Introduction

The complex nature of the anatomical structures involved in hand and wrist injuries means that assessment tools are crucial to providing accurate evaluation and effective treatment [1]. To ensure thorough evaluation, Patient Reported Outcomes Measures (PROMs) provide valuable insights into the potential risks or benefits of an intervention [2], influencing clinical decision-making and assessing the effectiveness of treatment [3].

The upper extremity is one of the areas where the most progress has been made in the use of specific questionnaires to assess outcomes [4]. Reliable and validated hand-injury questionnaires are now available, such as the Michigan Hand Outcomes Questionnaire (MHQ). Developed at the University of Michigan in 1998, the MHQ uses rigorous psychometric principles to provide a multidimensional measure of the health status of patients with hand and wrist disorder. It consists of 37 items assessing 6 domains: overall hand function, activities of daily living, work, pain, aesthetics and satisfaction. It assesses the left and right hand separately, avoiding dominance effect and providing two distinct scales for aesthetics and satisfaction [5]. The MHQ demonstrated reliability, validity and sensitivity to change in all types of hand and wrist conditions: carpal tunnel syndrome, distal radius fracture, wrist pain and digital contracture, Dupuytren's disease, and rheumatoid arthritis [6–10]. In addition, together with the Disability of Arm, Shoulder and Hand (DASH) questionnaire [11], it is the most widely used PROM for the hand and wrist function [12].

Hand use is influenced by the customs of different cultures, such as the use of cutlery in Western countries as opposed to chopsticks in countries such as Japan, or the predominance of the

* Corresponding author.
E-mail addresses: mv.martinez@mutuamontanesa.es (M.V. Martínez-Fernández), isandoval@ugr.es (I.S. Hernandez), jesus.martinez@ual.es (J. Martínez-Cal), carmen.sarabia@unican.es (C. Sarabia-Cobo).

https://doi.org/10.1016/j.hansur.2024.101715

right hand for grooming, eating or greeting in Arab countries [13]. This diversity of use influences the difficulty of the tasks to be performed, which is reflected in responses to the corresponding questionnaire items and therefore in their psychometric properties [14]. A well-validated and culturally adapted instrument allows better clinical and research application [15]. In this context, the MHQ has been translated into several languages and cultures [16] using cross-cultural adaptation processes to ensure its content and conceptual validity.

Systematic reviews of PROMs provide an efficient evidence-based method for selecting a particular PROM for clinical or research purposes. Systematic quality reviews can identify shortcomings in these tools, and provide the basis for further studies to optimize their properties [17]. Although there have been other reviews of the MHQ [18,19], they were not based on cross-cultural validation studies.

Due to the numerous cross-cultural adaptations and validations that have been carried out on the MHQ and the different methodologies used, we thought it would be interesting to conduct this systematic review. The main aim was to provide a comprehensive synthesis and in-depth analysis of the quality of psychometric properties in different cross-cultural versions of the MHQ. The specific aim was to identify existing gaps and provide evidence-based suggestions for improving future cross-cultural versions of the MHQ.

2. Materials and methods

2.1. Design and protocol

To conduct this systematic review, we followed the methodological criteria outlined in the current consensus-based standards for the selection of health measurement instruments, COSMIN [17,20], for systematic reviews. In addition, the general recommendations of the PRISMA statement were followed [21]. The study was registered in the PROSPERO database for systematic reviews (CRD n° 42022303561).

2.2. Sources and search

The search phase was initiated by consulting the MEDLINE (via PubMed), Web of Science, CINAHL and SCOPUS databases. The keywords were: "Michigan Hand Outcomes Questionnaire", "surveys", "questionnaires", "validation", "cross-cultural" and "version", combined with the Boolean operators "AND" / "OR". Articles published up to 2 September 2023 were included.

2.3. Selection criteria

The following inclusion criteria were applied: (1) cross-cultural adaptation and validation methodology of the MHQ; (2) sample of hand and wrist patients; (3) psychometric properties with total and/or domain scores of the MHQ; (4) without language restriction; and (5) sample population over 18 years of age. Exclusion criteria comprised: (1) studies using the MHQ as an outcome measure; (2) studies using the MHQ to validate another instrument; (3) studies with inconclusive results; and (4) studies using the short version of the MHQ.

2.4. Selection of documents

The retrieved documents were transferred to the Rayyan platform (rayyan.qcri.org) [22] where citation titles and Abstracts were collected, reviewed and evaluated. Duplicate articles were removed. Two researchers (MVMF and CSC) independently screened titles and abstracts, excluding articles that did not meet the inclusion criteria. Articles meeting the criteria were selected and located for full-text reading, including those that raised doubts or where the title and Abstract did not provide sufficient information for definitive inclusion or exclusion. Disagreements were resolved by a third reviewer (JMC).

2.5. Results, synthesis and data extraction

All finally selected articles were thoroughly analyzed to extract information on the construction and validation processes of these tools. Structural characteristics comprised: title, authors, year of publication, version, population, sample size, age, gender, setting, geographical location, target population, important disease characteristics, and number of subjects in the pilot phase and number of subjects per item in the validation phase (Table1). Psychometric properties were assessed according to the latest and improved version of the COSMIN guidelines [17,20,23]: test-retest reliability, internal consistency and measurement error and validity for structural validity and hypothesis testing for construct validity. Responsiveness was also included. Corresponding indices comprised intraclass correlation coefficient (ICC) for test-retest reliability, Cronbach's alpha for internal consistency (Table 2), standard deviation of change (SDC), limits of agreement (LoA), minimal important change (MIC), standard error of measurement (SEM) for measurement error, comparative fit index (CFI), Tucker-Lewis index (TLI), root mean square error of approximation (RMSEA) or standardized root mean square residual (SRMR) for structural validity, Spearman *rho* or Pearson *r* for construct validity (Table 2) and effect size (ES) or standardized response mean (SRM) for responsiveness.

2.6. Assessment of methodological quality

This assessment was developed following the risk-of-bias checklist [20,23]. It consists of a series of standards presented in "boxes", each of which assigns a rating to the quality of the measurement property in each study. The rating was given on a scale of "very good", "adequate", "doubtful" or "inadequate". The "worst score counts" principle was applied to the overall of all the scores for each property in the study to produce the final score [24].

For the purposes of hypothesis testing for construct validity, this assessment was made for each comparison instrument.

2.7. Criteria for good measurement properties

The results of each measurement property in each study were then assessed by consensus between two reviewers (MVMF and ISH) using the updated criteria for measurement properties. The results were classified as "+" *sufficient*, "−" *insufficient* or "?" *indeterminate*, depending on whether they met the COSMIN criteria, did not meet them or did not provide the required information. If a property was labelled "?" *indeterminate* or "±" *inconsistent*, no further analysis was performed and the quality of the evidence could not be assessed [17,20].

In this section, assessment also included construct validity and responsiveness. The hypotheses were developed by the review team to bring more uniformity to the evaluation [20]. Construct validity required $\geq$ 0.30 Pearson *r* or Spearman *rho* correlations between PROMs with the same or similar constructs, and responsiveness was based on ES or SRM considering medium to high values $\geq$ 0.5.

2.8. Synthesis of the evidence

Thirdly, a synthesis of the evidence was conducted by qualitatively assessing the measurement properties of each

Table 1
Structural characteristics of the questionnaires.

Questionnaire/author, year reference/ version	Population/sample size, age, gender	Setting/geographical location	Target population	Important disease characteristics	Number of subjects – pilot phase	Number of subjects per item
Validación del "Michigan Hand Outcomes Questionnaire" para población colombiana/Miranda et al. 2008/ Colombian-Spanish version [28]	N = 205 patients (122 male, 83 female) Median: 39 years old (range, 29.2–50)	Plastic Surgery, Rheumatology Plastic Surgery, Rheumatology and Rehabilitation of the University Hospital San Vicente de Paúl in Medellín (HUSVP), Antioquia, Colombia	Patients who consulted with symptoms with less than three months' evolution	Hand disorders, trauma, CTS and RA	20	5.54
Responsiveness of the Michigan Hand Outcomes Questionnaire- Dutch Language Version in Patients With Rheumatoid Arthritis/van der Giesen et al./2008/Dutch version [42]	N = 28 patients (4 male, 24 female) with an overall age of 54.0 years (range, 21–80)	Day patient clinic of the Leiden University Medical Center. Leiden, Netherlands	Patients with rheumatic disease and a complex hand function problems.	Presence of ulnar deviation, boutonnière and swan-neck deformities.	NR	0.75
Traducao e adaptacao transcultural do Michigan Hand Outcomes Questionnaire/ Ribeiro et al.2011 /Brazilian version [53]	N = 30 patients (15 male, 15 female)	NR	Rheumatoid arthritis	Rheumatoid arthritis	10	0.81
Cross-Cultural Adaptation, Validation, and Reliability Process of the Michigan Hand Outcomes Questionnaire in a Turkish Population/Öksüz et al. 2011/Turkish version [31]	N = 70 patients (28 male, 42 female) with an overall age of 42 ± 11 years (range, 21–56)	Outpatient department of physiotherapy or occupational therapy at Hacettepe University, Faculty of Health Science Samanpazari Ankara, Turkey	Different hand problems of at least 4 weeks' duration	Nerve injury, fractures, soft tissue problems (tendon injuries, ligament injuries), tendinitis, trigger finger other problems	30	1.89
Cross-cultural Adaptation and Validation of the Korean Version of the Michigan Hand Questionnaire/Roh et al. 2011/Korean version [38]	N = 176 patients 67 male 109 female Age 49 years (18–75)	Seoul National University Bundang Hospital, Seongnam; and National Police Hospital Seoul, Korea	Common hand problems	CTS Distal radius fracture Entrapment tenosynovitis Scaphoid fracture Benign soft tissue mass Osteoarthritis Other	10	4.75
Responsiveness of the Korean Version of the Michigan Hand Outcomes Questionnaire after Carpal Tunnel Release/Wi et al. 2013/Korean version [41]	N = 37 patients 1 male 36 female (overall age, 53.5 years)	An urban tertiary referral hospital Gachon University Hospital Incheon, Seoul, Korea	CTS	CTS scheduled for surgery	NR	1
Cross-cultural adaptation and validation of the Michigan Hand Outcomes Questionnaire (MHQ) for Brazil: validation study/Meireles et al. 2014/ Brazilian-Portuguese version [32]	N = 60 patients 30 patients adaptation/30 patients validation Age 49.9 ± 9.3 years Female 25 (83.4%)/Male 5 (16.6%)	Hospital Universitário Evangélico de Curitiba (HUEC), Curitiba, Paraná, Brazil	RA	RA	30	1.62
Measurement Properties of the German Michigan Hand Outcomes Questionnaire in Patients with Trapeziometacarpal Osteoarthritis/Marks et al. 2014/German version [35]	N = 177 patients Female 145 (82%) Age (years) 63.5 ± 9.2	Schulthess Klinik, Zurich, Switzerland, Ethikkommission Zurich, Switzerland	TMC joint OA diagnosed	TMC joint OA treatment consisted in conservative management (injection, analgesics, or occupational therapy) or surgery (resection/ suspension/interposition arthroplasty or arthrodesis)	NR	4.78
Cross-cultural adaptation, validation and reliability of the Michigan Hand Outcomes Questionnaire among Persian population/Ebrahimzadeh et al. 2015/ Persian version [15]	N = 223 patients Age, mean (SD) 35 years (15) Gender Male 114 (51%) Female 109 (49%)	Mashhad Iran	Common hand disorders	Trigger finger Non-specific pain Lunate osteonecrosis CTS Fracture Amputation 1 st CMC OA De Quervain Ganglion cyst Other	30	6.02

Table 1 (*Continued*)

Questionnaire/author, year reference/ version	Population/sample size, age, gender	Setting/geographical location	Target population	Important disease characteristics	Number of subjects – pilot phase	Number of subjects per item
Cultural adaptation of the Michigan Hand Outcomes Questionnaire in patients with Carpal Tunnel Syndrome: A Turkish version study/Ilhanli et al. 2015/Turkish version [40]	N = 150 Age (years) Mean ± SD 49.82 ± 9.54 42 male 108 female	NR	CTS	CTS	10	4.05
Reliability and Validity of the Japanese Version of the Michigan Hand Outcomes Questionnaire: A Comparison with the DASH and SF-36 Questionnaires/Oda et al. 2016/Japanese version [33]	N = 68 patients Age (years) Mean ± SD 57.7 ± 14.7 21 male 47 female	Sapporo Medical University, General Hospital, Shimonoseki Uji Takeda Hospital, Uji, Seirei Hamamatsu Hospital, Hamamatsu, International University of Health and Welfare, Tokyo, Showa University, Tokyo, Japan	Some hand disorders	CTS Trigger finger Cubital tunnel syndrome Osteoarthritis TFCC injury Malunion of fracture Ulnar impaction syndrome Chronic tendon problem De Quervain disease Chronic finger ligament injury Osteonecrosis Nerve palsy Dupuytren's disease Bone tumor	30	1.83
French translation and cross-cultural adaptation of the Michigan Hand Outcomes Questionnaire and the Brief Michigan Hand Outcomes Questionnaire/ Efanov et al. 2017/Canadian-French version [50]	N = 30 patients 14 male 16 female	University of Montreal Hospital Centre Montreal. Canada	Reasons for rehabilitation protocols	CTS Cubital tunnel syndrome OA Injury Dupuytren's disease Trigger finger Other	30	0.81 MHQ
The Polish version of the Michigan Hand Outcomes Questionnaire: Cross-cultural adaptation, reliability, construct validity and measurement error/Koziej et al. 2017/Polish version [36]	N = 120 patients Male 38 (32%) Female 82 (68%) Age 56.4 ± 14.9 years	Second Department of General Surgery, Jagiellonian University Medical College Krakow, Poland	Hand/wrist condition, no surgery on the affected hand/wrist in the last 6 months	CTS Cubital tunnel syndrome OA Injury Dupuytren's disease Trigger finger Other	11	3.24
Cross-Cultural Adaptation, Validity, and Reliability of the Patient-Rated Michigan Hand Outcomes Questionnaire for Thai Patients/Dhippayom et al. 2018/Thai version [34]	N = 30 Male 9 (30%) Female 21 (70%) Age, 51.86 ± 11.13 years	Occupational therapy units of four hospitals in three provinces: Chiangmai, Phitsanulok, and Bangkok. Thailand	Hand injury from an orthopedic condition	Bone injury: fracture Joint degeneration: arthritis and subluxation Muscle, tendon, and ligament injury Nerve injury Unknown	10	0.81
The Michigan Hand Questionnaire and Brief Michigan Hand Questionnaire were successfully translated to Canadian French/Busuioc et al. 2018/Canadian French version [30]	N = 78	Outpatient hand clinics at the Center Hospitalier de l'Université de Montréal Montréal, Québec, Canada	Diagnosed with any hand or wrist condition	Wide variety of hand conditions, representative of the population of patients usually seen in hand surgery and rehabilitation outpatient clinics	10	2.10
Cross-Cultural Adaptation and Validation of the Finnish Version of the Michigan Hand Outcomes Questionnaire/Hulkkonen et al. 2020/Finnish version [27]	N = 115 67 female 48 male	Department of Hand Surgery Helsinki University Hospital, Helsinki Finland	Various hand problems treated in general orthopedic outpatient clinic	CTS Trigger finger Ganglion cyst Dupuytren's disease 1 st CMC arthrosis Distal radius fracture Other fracture of the hand/wrist Other	10	3.10

Table 1 (*Continued*)

Questionnaire/author, year reference/ version	Population/sample size, age, gender	Setting/geographical location	Target population	Important disease characteristics	Number of subjects – pilot phase	Number of subjects per item
Assessment of construct validity of the Finnish versions of the Disabilities of Arm, Shoulder and Hand Instrument and the Michigan Hand Outcomes Questionnaire/Uimonen et al. 2020/ Finnish version [39]	N = 193 Age (years), 54 ± 15 Female: 114 (59%)	General orthopedic outpatient clinic Länsi-Pohja Central Hospital in Kemi, Finland	Heterogeneous sample of patients with hand and wrist problems	CTS Trigger finger Distal radius fracture Other fracture of the hand/wrist Ganglion cyst Dupuytren's disease CMC Joint OA Ulnar nerve entrapment Other	NR	5.21
A Thai version of the Michigan hand questionnaire (Thai MHQ): an investigation of the psychometric properties/Atthakomol et al. 2020/Thai version [26]	N = 217 Age (years); mean ± SD 47 ± 17 Female: 143 (66%)	Hand Outpatient Clinic of Academic University Hospital Chiang Mai, Thailand	Hand/wrist musculoskeletal disorders,	Joint stiffness Nerve injury Wrist fracture Tendon entrapment Compound fracture Joint arthritis Nerve entrapment Tumor tendon injury Hand infection Ligament injury Calcific tendinitis Joint dislocation Hand fracture Malunion Soft tissue injury Finger amputation	30	5.86
The Italian Version of the Michigan Hand Outcomes Questionnaire (MHQ): Translation, Cross-Cultural Adaptation and Validation/Passiatore et al. 2021/ Italian version [54]	N = 136 (48 male, 88 female) Age (years): mean ± SD 57 ± 15.7	Istituto di Clinica Ortopedica, Fondazione Policlinico Universitario A. Gemelli IRCCS, Università Cattolica del Sacro Cuore, Roma, Italy	Pathology involving one hand for at least 4 weeks	Carpal tunnel syndrome, osteoarthritis, tendonitis, Dupuytren's disease, ulnar nerve compression	55	3.67
The Malay Version of the Michigan Hand Outcomes Questionnaire: Cross-Cultural Adaptation, Validation and Reliability Testing/Lye et al. 2021/Malay version [48]	N = 100 (47 male, 53 female) Age (years): mean ± SD 49.43 ± 15.35	Hand Clinic at International Islamic University Malaysia Medical Centre (IIUMMC). Kuantan, Malaysia	Patients with various hand disorders	NR	30	2.70
The Michigan Hand Outcomes Questionnaire (MHQ-Swe) in patients with distal radius fractures—cross-cultural adaptation to Swedish, validation and reliability/Blomstrand et al. 2021/Swedish version [29]	N = 78 patients Age, years 61.1 ± 13.5 63 female (81%) 27 male (19%)	Orthopaedic Hand Therapy Unit at a university hospital in southwestern Sweden, Gothenburg, Sweden	A distal radius fracture	Distal radius fracture, all treated surgically using a volar plate	40	2.10
Cross-cultural adaptation & validity of the patient-rated Michigan Hand outcomes questionnaire (MHQ) in Bahasa Melayu for Malaysian patients/Kumar et al. 2022/ Malaysian version [37]	N= 185 91 female (49.2%) 94 male (50.8%)	Hospital Tengku Ampuan Afzan (HTAA), Hospital 159 Serdang (HSDG) and Hospital Pengajar Universiti Putra Malaysia (HPUPM)	Patients with hand and/ or wrist injuries	OA CRPS De Quervain's disease Fracture Ganglion Infection Nerve Compression & Injury Tendon Injury Trigger Finger/Thumb Soft Tissue Tumor	30	5

SD: Standard Deviation; NR: Not Reported; CTS: Carpal Tunnel Syndrome; RA: Rheumatoid Arthritis; TMC: Trapeziometacarpal; OA: Osteoarthritis; CMC: Carpometacarpal; TFCC: Triangular FibroCartilage Complex; CRPS: Complex Regional Pain Syndrome.

Table 2
Reliability, internal consistency and construct validity scores.

Study/version	Test-retest Reliability	Internal consistency	Construct validity – convergent validity
Miranda et al. 2008/ Colombian-Spanish version [28]	ICC (95% CI) D1: 0.66 (0.45–0.79) D2: 0.83 (0.71–0.90) D3: 0.63 (0.42–0.77) D4: 0.48 (0.22–0.67) D5: 0.75 (0.60–0.88) D6: 0.77 (0.67–0.87)	Cronbach's α D1: 0.92 D2: 0.92 D3: 0.96 D4: 0.91 D5: 0.78 D6: 0.85 Total: 0.92	MHQ - SF-36 Physical function: (*rho*) Spearman D1: 0.46; D2: 0.57; D3: 0.33; D4: 0.33; D5:0.37; D6: 0.48 MHQ - SF-36 Role physical: (*rho*) Spearman D1: 0.05; D2: 0.35; D3: 0.48; D4: -0.004; D5: 0.13; D6: 0.04 MHQ - SF-36 Pain: (*rho*) Spearman D1: 0.33; D2: 0.22; D3: -0.17; D4: 0.78; D5: 0.43; D6: 0.50
van der Giesen et al. 2008/ Dutch version [42]	NR	NR	NR
Ribeiro et al. 2011./ Brazilian version [53]	NR	NR	NR
Öksüz et al. 2011/Turkish version [31]	ICC (R/ L)/(R/ L/ B) D1: (0.82/0.95) D2: (0.89/0.92/0.91) D3: 0.79 D4: 0.91 D5: (0.95/0.96) D6: (0.91/0.94)	Cronbach's α (R/ L)/ (R/ L/ B) D1: (0.88/ 0.94) D2: (0.96/ 0.95/ 0.87) D3: 0.94 D4: 0.85 D5: (0.76/ 0.79) D6: (0.94/ 0.96)	MHQ - DASH Function/Symptom: (*rho*) Spearman D1: -0.46; D2: -0.66; D3: -0.48; D4: 0.48; D5: -0.33; D6: -0.44 MHQ - DASH Sports/Music: (*rho*) Spearman D1: -0.07; D2: 0.03; D3: -0.01; D4: 0.27; D5: 0.04; D6: -0.13 MHQ - DASH Work: (*rho*) Spearman D1: -0.18; D2: -0.36; D3: -0.40; D4: 0.27; D5: -0.24; D6: -0.22 MHQ - VAS: (*rho*) Spearman D1: -0.47; D2: -0.31; D3: -0.23; D4: 0.33; D5: -0.12; D6: -0.32 MHQ - Grip Injured: (*rho*) Spearman D1: 0.32; D2: 0.52; D3: 0.34; D4: -0.13; D5: 0.35; D6: 0.30
Roh et al. 2011/Korean version [38]	ICC (R/ L)/ (R/ L/ B) D1: (0.93/ 0.92) D2: (0.95/ 0.93/ 0.91) D3: 0.96 D4: 0.93 D5: (0.88/ 0.90) D6: (0.95/ 0.94) Total ICC:0.94	Cronbach's α = 0.84 (R/ L)/ (R/ L/ B) D1: (0.95/ 0.96) D2: (0.96/ 0.97/ 0.85) D3: 0.97 D4: 0.85 D5: (0.79/ 0.80) D6: (0.96/ 0.96) Total:0.84	MHQ - DASH Disability/Symptom: (*rho*) Spearman D1: -0.46; D2: -0.59; D3: -0.61; D4: 0.63; D5: -0.32; D6: -0.48 MHQ - DASH Work: (*rho*) Spearman D1: -0.34; D2: -0.26; D3: -0.38; D4: 0.40; D5: -0.18; D6: -0.17 MHQ - DASH Sport/Music: (*rho*) Spearman D1: 0.04; D2: 0.05; D3: 0.00; D4: -0.05; D5: 0.00; D6: 0.20 MHQ - Pain VAS: (*rho*) Spearman D1: 0.09; D2: 0.32; D3: 0.37; D4: -0.60; D5: 0.12; D6: 0.27 MHQ - Functional State: (*rho*) Spearman D1: 0.07; D2: 0.31; D3: 0.34; D4: -0.56; D5: 0.10; D6: 0.28 MHQ - Range of motion: (*rho*) Spearman D1: 0.22; D2: 0.27; D3: 0.27; D4: -0.58; D5: -0.25; D6: 0.06 MHQ - Grip power: (*rho*) Spearman D1: 0.20; D2: 0.37; D3: 0.35; D4: -0.59; D5: -0.12; D6: 0.01
Wi et al./Korean version [41]	NR	NR	NR
Meireles et al. 2014/ Brazilian-Portuguese version [32]	ICC (R/ L)/(R/ L/ B) D1: IRE 1: (0.915/0.908); IRE 2: (0.863/ 0.875) D2: IRE 1: (0.901/ 0.841/0.967); IRE 2: (0.611/ 0.783/0.818) D3: IRE 1: 0.918; IRE 2: 0.753 D4: IRE 1: (0.929/0.944); IRE 2: (0.885/ 0.826) D5: IRE 1: (0.929/0.919); IRE 2: (0.921/0.905) D6: IRE 1: (0.883/0.937); IRE 2: (0.786/0.876) Total Right: IRE 1 (0.976); IRE 2 (0.917) Total Left: IRE 1 (0.980); IRE 2 (0.936)	Cronbach's α (R/ L)/ (R/ L/ B) D1: (0.908/ 0.941) D2: (0.871/ 0.939/ 0.930) D3: 0.969 D4: (0.908/ 0.941) D5: (0.864/ 0.925) D6: (0.940/ 0.939)	MHQ - HAQ: (*rho*) Spearman D1: -0.62; D2: -0.64; D3: -0.72; D4: 0.62; D5: -0.44; D6: -0.47 MHQ - DASH: (*rho*) Spearman D1: -0.67; D2: -0.70; D3: -0.69; D4: 0.65; D5: -0.41; D6: -0.55 MHQ - DASH Work: (*rho*) Spearman D1: -0.73; D2: -0.59; D3: -0.79; D4: 0.62; D5: -0.43; D6: -0.64 MHQ - Cochin: (*rho*) Spearman D1: -0.56; D2: -0.81; D3: -0.59; D4: 0.51; D5: -0.50; D6: -0.48 MHQ - VAS: (*rho*) Spearman D1: -0.60; D2: -0.43; D3: -0.50; D4: 0.75; D5: -0.52; D6: -0.58
Marks et al. 2014/German version [35]	ICC D1: 0.85 D2: 0.89 D3: 0.94 D4: 0.92 D5: 0.85 D6: 0.88 Total score 0.95	Cronbach's α D1: 0.81 D2: 0.89 D3: 0.87 D4: 0.77 D5: 0.86 D6: 0.84	MHQ - DASH: (*r*) Pearson D1: -0.44; D2: -0.76; D3: -0.65; D4: 0.67; D5: -0.30; D6: -0.64 MHQ - SF-12 Physical: (*r*) Pearson D1: 0.23; D2: 0.45; D3: 0.56; D4: -0.46; D5: 0.12; D6: 0.38 MHQ - SF-12 Mental: (*r*) Pearson D1: 0.21; D2: 0.30; D3: 0.39; D4: -0.30; D5: 0.32; D6: 0.35 MHQ - Key Pinch (Kg): (*r*) Pearson D1: 0.36; D2: 0.44; D3: 0.25; D4: -0.32; D5: 0.31; D6: 0.41
Ebrahimzadeh et al. 2015/ Persian version [15]	ICC (95% CI) D1: 0.81 (0.32–0.94) D2: 0.78 (0.41–0.92) D3: 0.86 (0.64–0.95) D4: 0.78 (0.65–0.86) D5: 0.84 (0.75–0.90) D6: 0.73 (0.31–0.90) Total: 0.84 (0.60–0.94)	Cronbach's α (T1/ T2) D1: 0.65/ 0.80 D2: 0.96/ 0.92 D3: 0.92/ 0.94 D4: 0.79/ 0.87 D5: 0.83/ 0.92 D6: 0.83/ 0.63 Total: 0.79/ 0.75	MHQ - DASH: (*rho*) Spearman D1: -0.70; D2: -0.75; D3: -0.53; D4: 0.41; D5: -0.24; D6: -0.67 MHQ - DASH Work: (*rho*) Spearman D1: -0.60; D2: -0.52; D3: -0.39; D4: 0.24; D5: -0.21; D6: -0.64 MHQ - DASH Sport: (*rho*) Spearman D1: -0.48; D2: -0.55; D3: -0.47; D4: 0.24; D5: -0.38; D6: -0.61 MHQ - VAS: (*rho*) Spearman D1: -0.23; D2: -0.21; D3: -0.10; D4: 0.15; D5: 0.03; D6: -0.22

Table 2 (*Continued*)

Study/version	Test-retest Reliability	Internal consistency	Construct validity – convergent validity
Ilhanli et al. 2015/Turkish version [40]	ICC (R/ L)/ (R/ L/ B) D1: (0.99/ 0.99) D2: (0.99/0.99/ 0.98) D3: 0.98 D4: 0.99 D5: (0.99/ 0.99) D6: (0.99/0.99)	Cronbach's α (R/ L)/ (R/ L/ B) D1: (0.96/ 0.97) D2: (0.94/ 0.95/ 0.94) D3: 0.97 D4: 0.96 D5: (0.92/ 0.92) D6: (0.97/ 0.97)	MHQ-D1, D2, D3, D4, D5, D6: MHQ -Function; (*rho*) Spearman D2:0.77; D3:0.88; D4: -0.60; D5:0.71; D6:0.20 MHQ – ADL;*(rho)* Spearman D3:0.74; D4: -0.65; D5:0.73; D6:0.19; Total ADL:0.955 MHQ – Work; (rho)Spearman D4: -0.59; D5:0.65; D6:0.31 MHQ – Pain; (*rho*) Spearman D5: -0.66; D6: -0.13 MHQ -Aesthetics(rho) Spearman D6: 0.34
Oda et al. 2016/Japanese version [33]	ICC (R/ L)/ (R/ L/ B) D1: (0.892/ 0.906) D2: (0.924/0.899/ 0.902) D3: 0.880 D4: (0.867/ 0.899) D5: (0.814/ 0.683) D6: (0.933/0.908) Total (95% CI): 0.997 (0.994–0.998)	Cronbach's α (R/ L)/ (R/ L/ B) D1: (0.944/ 0.951) D2: (0.962/ 0.948/ 0.954) D3: 0.936 D4: (0.928/ 0.946) D5: (0.896/ 0.809) D6: (0.965/ 0.951)	MHQ - DASH: (*rho*) Spearman D1: -0.49; D2: -0.81; D3: -0.81; D4: 0.65; D5: -0.52; D6: -0.65; Total-0.82 MHQ - SF-36 Physical function: (*rho*) Spearman D1: 0.34; D2: 0.51; D3: 0.53; D4: -0.47; D5: 0.34; D6: 0.33 MHQ - SF-36 Role physical: (*rho*) Spearman D1: 0.37; D2: 0.55; D3: 0.72; D4: -0.53; D5: 0.51; D6: 0.51 MHQ - SF-36 Body pain: (*rho*) Spearman D1: 0.40; D2: 0.48; D3: 0.61; D4: -0.58; D5: 0.33; D6: 0.50 MHQ - SF-36 General health perception: (*rho*) Spearman D1: 0.34; D2: 0.26; D3: 0.31; D4: -0.21; D5: 0.48; D6: 0.42 MHQ - SF-36 Vitality: (*rho*) Spearman D1: 0.26; D2: 0.40; D3: 0.38; D4: -0.25; D5: 0.44; D6: 0.32 MHQ - SF-36 Social functioning: (*rho*) Spearman D1: 0.31; D2: 0.58; D3: 0.44; D4: -0.36; D5: 0.48; D6: 0.35 MHQ - SF-36 Role emotional: (*rho*) Spearman D1: 0.24; D2: 0.54; D3: 0.56; D4: -0.48; D5: 0.50; D6: 0.41 MHQ - SF-36 Mental health: (*rho*) Spearman D1: 0.22; D2: 0.45; D3: 0.41; D4: -0.42; D5: 0.45; D6: 0.39
Efanov et al. 2017/Canadian French version [50]	NR	NR	NR
Koziej et al. 2017/Polish version [36]	ICC (R/ L)/ (R/ L/ B) D1: (0.93//0.95) D2: (0.95/ 0.97/ 0.94) D3: (0.88) D4: (0.91/ 0.89) D5: (0.82/ 0.83) D6: (0.93/ 0.94) Total Right: 0.95 Total Left: 0.96 Injury/dysfunctional hand:0.91	Cronbach's α (R/ L)/ (R/ L/ B) D1: (0.79/ 0.94) D2: (0.95/ 0.96/ 0.94) D3: 0.91 D4: (0.78/ 0.79) D5: (0.86/ 0.86) D6: (0.94/ 0.95)	MHQ - DASH Function/Symptoms: (*rho*) Spearman D1: -0.58; D2: -0.70; D3: -0.55; D4: 0.47; D5: -0.41; D6: -0.56 MHQ - DASH Work: (*rho*) Spearman D1: -0.51; D2: -0.42; D3: -0.51; D4: 0.41; D5: -0.36; D6: -0.47 MHQ - VAS: (*rho*) Spearman D1: -0.22; D2: -0.28; D3: -0.11; D4: 0.56; D5: -0.34; D6: -0.37 MHQ - Grip Test: (*rho*) Spearman D1: 0.33; D2: 0.37; D3: 0.38; D4: -0.17; D5: 0.17; D6: 0.21 MHQ - Pinch Test: (*rho*) Spearman D1: 0.20; D2: 0.26; D3: -0.05; D4: -0.22; D5: 0.19; D6: 0.29
Dhippayom et al. 2018/Thai version [34]	ICC (95% CI) D1: 0.912 (0.824–0.957) D2: 0.882 (0.767–0.942) D3: 0.776 (0.581–0.887) D4: 0.941 (0.881–0.972) D5: 0.799 (0.619–0.899) D6: 0.963 (0.924–0.982) Total: 0.953 (0.904–0.977)	Cronbach's α D1: 0.778 D2: 0.992 D3: 0.892 D4: 0.947 D5: 0.477 D6: 0.850 Total: 0.835	MHQ-D1, D2, D3, D4, D5, D6: MHQ -Function; (*rho*) Spearman D2:0.671; D3:0.369; D4: -0.693; D5:0.403; D6:0.807 MHQ – ADL;*(rho)* Spearman D3:0.706; D4: -0.731; D5:0.694; D6:0.743 MHQ – Work; (rho)Spearman D4: -0.605; D5:0.428; D6:0.648 MHQ – Pain; (*rho*) Spearman D5: -0.355; D6: -0.798 MHQ -Aesthetics;(rho) Spearman D6: 0.507
Busuioc et al. 2018/ Canadian-French version [30]	ICC (R/ L)/ (R/ L/ B) D1: (0.87/ 0.88) D2: (0.88/ 0.90/ 0.90) D3: 0.84 D4: (0.85/ 0.84) D5: (0.81/ 0.81) D6: (0.89/ 0.93) Total: (0.95/ 0.95) bMHQ: 0.86	Cronbach's α (R/ L)/ (R/ L/ B) D1: (0.87/ 0.90) D2: (0.97/ 0.95/ 0.92) D3: 0.95 D4: (0.87/ 0.90) D5: (0.83/ 0.79) D6: (0.93/ 0.94) bMHQ: 0.89	MHQ - DASH Disability symptom or Quick DASH: (*rho*) Spearman D1: -0.70; D2: -0.80; D3: -0.77; D4: 0.75; D5: -0.39; D6: -0.70 MHQ - DASH Work: (*rho*) Spearman D1: -0.72; D2: -0.86; D3: -0.87; D4: 0.71; D5: -0.33; D6: -0.71 MHQ - DASH Sport/Music: (*rho*) Spearman D1: -0.56; D2: -0.59; D3: -0.64; D4: 0.53; D5: -0.28; D6: -0.53 MHQ - SF-12 Physical component: (*rho*) Spearman D1: 0.51; D2: 0.56; D3: 0.64; D4: 0.59; D5: 0.16; D6: 0.55 MHQ - SF-12 Mental component: (*rho*) Spearman D1: 0.35; D2: 0.28; D3: 0.37; D4: 0.30; D5: 0.45; D6: 0.34 MHQ - Pain VAS: (*rho*) Spearman D1: -0.61; D2: -0.51; D3: -0.57; D4: 0.78; D5: -0.38; D6: -0.65
Hulkkonen et al. 2018/ Finnish version [27]	ICC (95% CI) (R/ L)/ (R/ L/ B) D1: 0.90 (0.84–0.94)/ 0.83 (0.74–0.90) D2: 0.89 (0.82–0.93)/ 0.77 (0.64–0.86)/ 0.76 (0.63–0.85) D3: 0.87 (0.79–0.92) D4: 0.88 (0.80–0.93)/ 0.82 (0.72–0.89) D5: 0.71 (0.55–0.81)/ 0.66 (0.50–0.78) D6: 0.91 (0.85–0.94)/ 0.86 (0.78–0.91)	Cronbach's α (R/ L)/ (R/ L/ B) D1: (0.899/ 0.929) D2: (0.925/ 0.972/ 0.936) D3: 0.941 D4: (0.934/ 0.942) D5: (0.799/ 0.861) D6: (0.914/ 0.939)	MHQ - DASH: (*rho*) Spearman D1: -0.72; D2: -0.80; D3: -0.75; D4: 0.71; D5: -0.45; D6: -0.68 MHQ - DASH Work: (*rho*) Spearman D1: -0.61; D2: -0.67; D3: -0.79; D4: 0.50; D5: -0.22; D6: -0.50 MHQ - DASH Sport/Music: (*rho*) Spearman D1: -0.54; D2: -0.55; D3: -0.68; D4: 0.59; D5: -0.24; D6: -0.48 MHQ - EQ 5-D: (*rho*) Spearman D1: 0.45; D2: 0.53; D3: 0.42; D4: -0.62; D5: 0.29; D6: 0.49 MHQ - EQ VAS: (*rho*) Spearman D1: 0.38; D2: 0.44; D3: 0.40; D4: -0.49; D5: 0.23; D6: 0.45 MHQ - Pain VAS: (*rho*) Spearman D1: -0.51; D2: -0.41; D3: -0.41; D4: 0.74; D5: -0.25; D6: -0.64 MHQ - Grip Strength (Kg): (*rho*) Spearman D1: 0.52; D2: 0.64; D3: 0.51; D4: -0.35; D5: 0.41; D6: 0.41 MHQ - Key Pinch Strength (Kg): (*rho*) Spearman D1: 0.47; D2: 0.55; D3: 0.38; D4: -0.31; D5: 0.33; D6: 0.42

Table 2 (*Continued*)

Study/version	Test-retest Reliability	Internal consistency	Construct validity – convergent validity
Uimonen et al. 2020/ Finnish version [39]	NR	NR	MHQ - EQ-5D Index: (*r*) Pearson (95% CI): 0.43 (0.28–0.56) MHQ - EQ-5D Mobility: (*r*) Pearson (95% CI): 0.21 (0.06–0.36) MHQ - EQ-5D Self-care: (*r*) Pearson (95% CI): 0.40 (0.23–0.53) MHQ - EQ-5D Usual activity: (*r*) Pearson (95% CI): 0.48 (0.34–0.60) MHQ - EQ-5D Pain, discomfort: (*r*) Pearson (95% CI): 0.25 (0.07–0.41) MHQ - EQ-5D Anxiety/Depression: (*r*) Pearson (95% CI):0.34 (0.17–0.47) MHQ - EQ-VAS: (*r*) Pearson (95% CI): 0.44 (0.28–0.57) MHQ - Pain VAS: (*r*) Pearson (95% CI): 0.34 (0.14–0.50)
Atthakomol et al. 2020/Thai version [26]	ICC (95% CI) D1: 0.93 (0.88–0.95) D2: 0.87 (0.79–0.92) D3: 0.83 (0.73–0.90) D4: 0.95 (0.92–0.97) D5: 0.87 (0.79–0.92) D6: 0.90 (0.84–0.94) Total: 0.95 (0.91–0.97)	Cronbach's α (R/ L)/ (R/ L/ B) D1: (0.94/ 0.94) D2: (0.97/ 0.97/ 0.96) D3: 0.94 D4: (0.97/ 0.98) D5: (0.79/ 0.80) D6: (0.95/ 0.96)	MHQ - DASH Common activities: (*rho*) Spearman D1: -0.58; D2: -0.77; D3: -0.58; D4: 0.57; D5: -0.38; D6: -0.59 MHQ - DASH Self-care activities: (*rho*) Spearman D1: -0.53; D2: -0.72; D3: -0.49; D4: 0.52; D5: -0.33; D6: -0.50 MHQ - DASH Pain: (*rho*) Spearman D1: -0.40; D2: -0.49; D3: -0.44; D4: 0.58; D5: -0.20; D6: -0.51 MHQ - DASH Other symptoms: (*rho*) Spearman D1: -0.33; D2: -0.41; D3: -0.35; D4: -0.54; D5: -0.32; D6: -0.47 MHQ - DASH Psychological effects: (*rho*) Spearman D1: -0.33; D2: -0.38; D3: -0.32; D4: 0.43; D5: -0.37; D6: -0.37 MHQ - EQ-5D-5 L Mobility: (*rho*) Spearman D1: -0.12; D2: -0.19; D3: -0.15; D4: 0.11; D5: -0.13; D6: -0.09 MHQ - EQ-5D-5 L Usual activities: (*rho*) Spearman D1: -0.44; D2: -0.53; D3: -0.47; D4: 0.48; D5: -0.30; D6: -0.50 MHQ - EQ-5D-5 L Self-care activities: (*rho*) Spearman D1: -0.37; D2: -0.55; D3: -0.33; D4: 0.35; D5: -0.31; D6: -0.34 MHQ - EQ-5D-5 L Pain: (*rho*) Spearman D1: -0.43; D2: -0.50; D3: -0.41; D4: 0.59; D5: -0.28; D6: -0.53 MHQ - EQ-5D-5 L Anxiety/Depression: (*rho*) Spearman D1: -0.20; D2: -0.33; D3: -0.25; D4: 0.33; D5: -0.30; D6: -0.38
Passiatore et al. 2021/ Italian version [54]	ICC (95% CI) (R/ L)/ (R/ L/ B) D1: 0.98 (0.96–0.99)/ 0.94 (0.89–0.97) D2: 0.96 (0.92–0.98)/ 0.96 (0.91–0.98)/ 0.93 (0.86–0.97) D3: 0.91 (0.82–0.96) D4: 0.93 (0.86–0.97) D5: 0.87 (0.75–0.94)/ 0.83 (0.68–0.92) D6: 0.98 (0.96–0.99)/ 0.94 (0.87–0.97)	Cronbach's α (R/ L)/ (R/ L/ B) D1: (0.99/ 0.97) D2: (0.98/ 0.98/ 0.96) D3: 0.95 D4: 0.96 D5: (0.93/ 0.91) D6: (0.99/ 0.97)	MHQ - DASH: (*rho*) Spearman D1: -0.52; D2: -0.65; D3: -0.62; D4: 0.52; D5: -0.41; D6: -0.52 MHQ - DASH Work: (*rho*) Spearman D1: -0.49; D2: -0.36; D3: -0.47; D4: 0.42; D5: -0.35; D6: -0.46 MHQ – VAS Pain: (*rho*) Spearman D1: -0.50; D2: -0.35; D3: -0.19; D4: 0.39; D5: -0.15; D6: -0.40 MHQ - Grip Strength: (*rho*) Spearman D1: 0.35; D2: 0.65; D3: 0.41; D4: -0.22; D5: 0.39; D6: 0.31
Lye et al. 2021/Malay version [48]	ICC (R/ L)/ (R/ L/ B) D1: 0.957/ 0.981 D2: 0.984/ 0.982/ 0.970 D3: 0.928 D4: 0.961 D5: 0.982/ 0.982 D6: 0.958/ 0.925	Cronbach's α (R/ L)/ (R/ L/ B) D1: (0.94/ 0.94) D2: (0.97/ 0.97/ 0.96) D3: 0.94 D4: (0.97/ 0.98) D5: (0.79/ 0.80) D6: (0.95/ 0.96)	MHQ - DASH: (*rho*) Spearman D1: -0.662; D2: -0.757; D3: -0.559; D4: 0.574; D5: -0.513; D6: -0.574
Blomstrand et al. 2021/ Swedish version [29]	ICC (95% CI) D1:0.83 (0.75–0.89) D2: 0.87 (0.78–0.93) D3: 0.84 (0.72–0.91) D4: 0.86 (0.78–0.91) D5: 0.77 (0.63–0.85) D6: 0.90 (0.84–0.94) Total: 0.92 (0.87–0.95)	Cronbach's α (T1/ T2) D1: (0.77/ 0.81) D2: (0.93/ 0.94) D3: (0.94/ 0.96) D4: (0.85/ 0.87) D5:(0.79/ 0.82) D6: (0.89/ 0.87)	MHQ - VAS: (*rho*) Spearman D1: -0.425; D2: -0.433; D3: -0.402; D4: -0.674; D5: -0.312; D6: -0.500 MHQ - Patient Rated Wrist Evaluation (PRWE) – Pain: (*rho*) Spearman D1: -0.496; D2: -0.551; D3: -0.453; D4: -0.737; D5: -0.417; D6: -0.568 MHQ - PRWE – Activity: (*rho*) Spearman D1: -0.546; D2: -0.786; D3: -0.557; D4: -0.550; D5: -0.402; D6: -0.641
Kumar et al. 2022/ Malaysian version [37]	ICC (95% CI) D1: 0.975 (0.956–0.986) D2: 0.960 (0.930–0.977) D3: 0.984 (0.972–0.991) D4: 0.984 (0.972–0.991) D5: 0.979 (0.963–0.988) D6: 0.974 (0.954–0.985) Total: 0.984 (0.971–0.991)	Cronbach's α D1: 0.884 D2: 0.958 D3: 0.966 D4: 0.593 D5: 0.242 D6: 0.934 Total 0.821	MHQ-D1, D2, D3, D4, D5, D6 MHQ -Function; (*rho*) Spearman D2:0.745; D3:0.703; D4: -0.744; D5:0.479; D6:0.804; Total ADL:0.742; Total Score: 0.870 MHQ – ADL;*(rho)* Spearman D3:0.826; D4: -0.617; D5:0.501; D6:0.665; Total ADL:0.955; Total Score:0.819 MHQ – Work; (rho)Spearman D4: -0.724; D5:0.591; D6:0.716; Total ADL:0.751; Total Score: 0.885 MHQ – Pain; (*rho*) Spearman D5: -0.474; D6: -0.766; Total ADL: -0.636; Total Score: -0.855 MHQ -Aesthetics;(rho) Spearman D6: 0.566; Total ADL:0.531; Total Score: 0.716 MHQ – Satisfaction;(rho) Spearman Total ADL: 0.669; Total Score:0.883

NR: not reported; D1: OHF (Overall Hand Function); D2: ADLs (Activities of Daily Living); D3: Work performance; D4: Pain; D5: Aesthetics; D6: Satisfaction; ICC: Intraclass Correlation Coefficient; CI: Confidence Interval; R: right hand; L: left hand; B: Both hands; IRE 1: Inter-rater evaluation; IRE 2: Intra-rater evaluation; bMHQ: brief MHQ; T1: test; T2: re-test; HAQ: Health Assessment Questionnaire.

version. Consistency was rated as "+" *sufficient* or "−" *insufficient* depending on whether ≥ 75% were rated "+" *sufficient* or "−" *insufficient*. The results from each version were then pooled and qualitatively assessed according to the criteria for good measurement properties. If the results were "±" *inconsistent,* either an explanation could be provided or else they remained "±" *inconsistent*. For results classified as "?" *indeterminate* or "±" *inconsistent*, the following step was not implemented. Results per measurement property and per version were presented, together with the quality of the evidence.

Quality of evidence was graded "high", "moderate", "low" or "very low" using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach [17,20]. Factors comprised risk of bias due to methodological quality, imprecision (i.e., sample size), inconsistency of results, and indirectness of evidence from a population other than the population of interest for the review.

Assessment was carried out by two reviewers, who reached consensus (MVMF and CSC); agreement was assessed on the recommendations of Mokkink et al. [25] for inter-rater agreement with valid values of >80%. The percentage was calculated by dividing the number of concordant reviewer ratings by the total number of ratings.

3. Results

Of the 493 documents retrieved from the databases, 307 were excluded as duplicates. Of the 186 remaining studies, 151 were excluded after title and Abstract screening. The remaining 35 studies were reviewed in full text, after which 13 were excluded: 8 because they were not cross-cultural studies, 2 because they did not have a validation phase, and 3 for various reasons such as not providing results. Finally, 22 studies were included in this systematic review, 20 of which 20 were analyzed for their psychometric properties. The entire selection process is shown in the flowchart (Fig. 1).

Percentage inter-rater agreement was 83%, considered adequate (Table 1).

3.1. Synthesis of quality of evidence

Methodological quality was assessed according to risk of bias on the "worst score counts" principle. The criteria for good measurement properties were then analyzed for each study. For each PROM, qualitatively summarized results per measurement property were presented, together with an overall quality criterion indicating good measurement properties. In addition, the quality of evidence was assessed using the GRADE approach.

3.2. Structural validity

In terms of methodological quality, two versions of the MHQ were rated as respectively *doubtful* [26] and *inadequate* [27] because they did not provide rotation methods or did not reach the required sample size. According to the criteria for good measurement properties, the Colombian Spanish version [28] was rated as "?" *indeterminate* for the criteria of classical test theory (CTT), and the Finnish version [27] was rated "−" *insufficient* because CFI and TLI indices below 0.9 were not reported. Quality of evidence could not be assessed for the Colombian version [28], the Thai version

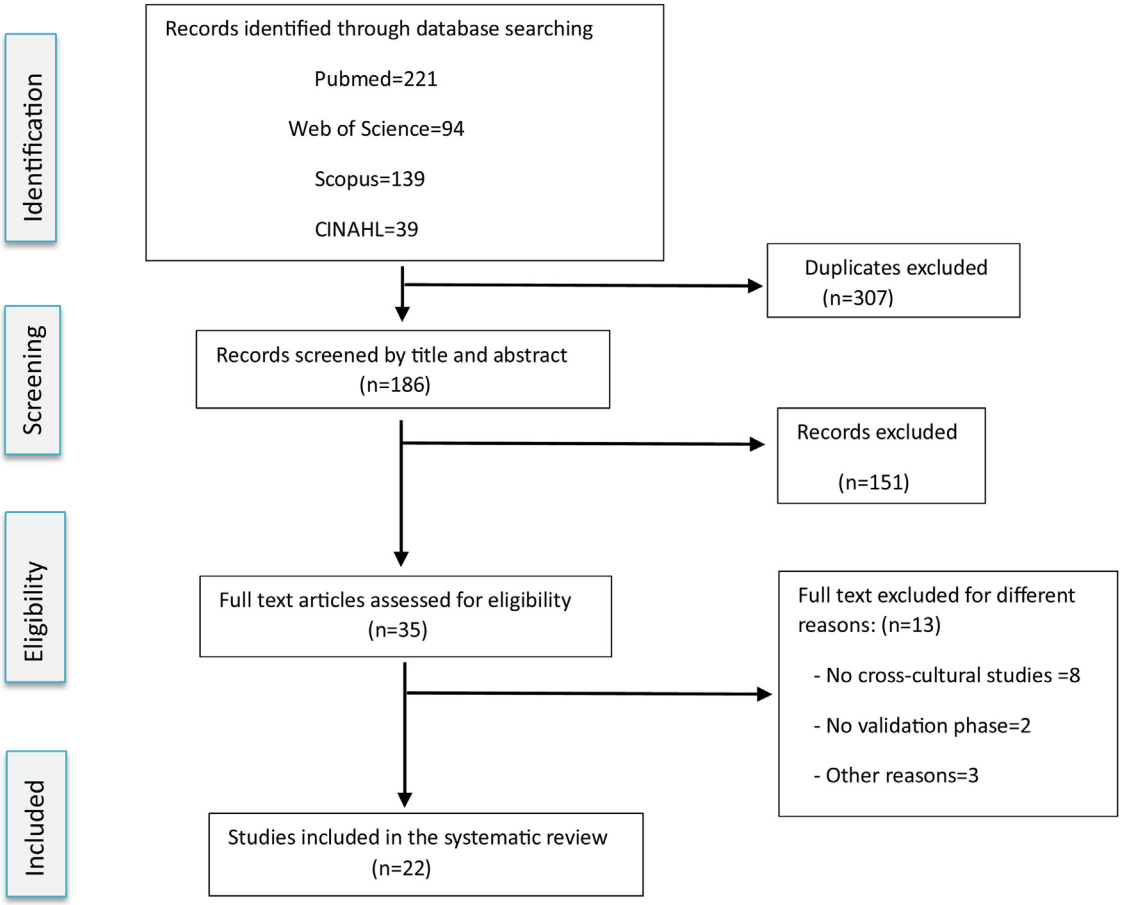

Fig. 1. Flowchart for selecting studies based on PRISMA.

[26] was rated *moderate* and the Finnish version [27] as *low* due to the imprecise sample size.

3.3. Internal consistency

Only three versions [26–28] could be assessed on structural validity analysis. Methodological quality was assessed by calculation of unidimensionality, Cronbach's alpha for each subscale and for continuous and dichotomous scores. Most versions were rated *doubtful,* because the calculation of unidimensionality was unclear, and two versions [26,28] were rated *very good*. The Colombian [28], Finnish [27] and Thai [26] versions [26,28] met the criteria for good measurement properties, rated "+" *sufficient* in both assessments. Quality of evidence was rated *high* in two versions [26,28] (Table 2).

3.4. Test-retest reliability

On analysis of methodological quality, three versions were rated *inadequate* because of sample instability or non-similarity in the test conditions [29–31] and the others as *doubtful* because of a time interval less than 2 weeks. With regard to good measurement properties, they were rated "+" *sufficient* in all versions, with ICC $\geq$ 0.70 in all domains, except in 4 versions [27,28,32,33], one of which was finally rated "±" *inconsistent* [28]. The quality of the evidence was rated *moderate* in all versions except for one in which it was rated *low* [34].

3.5. Measurement error

Methodological quality was considered *doubtful* in all three versions [26,35,36], due to short time intervals [35,36], instability, dissimilar conditions between two measurements [26] and for the calculation of SEM, SDC or LoA. The Thai [26] and German versions [26,35] were rated "-" *insufficient;* the latter had an SDC lower than the MIC in five out of seven results, thus failing to reach 75% of the required results. The Polish version [36] was rated "?" *indeterminate*. Quality of evidence was *moderate* in the Thai version [26] and *low* in the German version [35], partly due to sample size <50 patients.

3.6. Hypothesis testing for construct validity

This psychometric property assessed the methodological quality of each comparator instrument and was found to be *adequate* in most cases, *inadequate* in five versions [15,33,34,37,38] and *very good* in three [32,35,39]. This assessment was based on the clarity of the comparator construct, the measurement properties of the comparator PROMs in a similar population, and the appropriateness of the statistical methods used. Measurement properties were generally rated "+" *sufficient* in the 13 versions analyzed, and "?" *indeterminate* in three versions [34,37,40]. Subsequently, quality of the evidence was rated *moderate*, except for three versions [31,35,39] which were *high* and three which could not be rated due to a previous rating of "?" *indeterminate* [34,37,40].

3.7. Responsiveness

The assessment of the risk of bias of this methodological property took into account the hypotheses proposed by the review team, with four of the versions rated *very good* [26,35,41,42] and one *doubtful* [28]. Responsiveness as measurement property was rated "+" *sufficient* in three versions [26,35,41] because more than 75% of the effect size or SRM results were above 0.5 or the mean effect size, which is the overalleffect size set for the review [43], while two versions were rated "-" *insufficient* and "?" *indeterminate*, respectively [28,42]. Finally, quality of evidence was rated *moderate* in three versions [26,35,41] and *low* in one [42].

The results for methodological quality and quality of evidence and the criteria for good measurement properties are presented in Tables 3 and 4.

4. Discussion

This review provided, for the first time, an overview of the quality of the various cross-cultural versions of the MHQ. It included 22 papers, representing the 20 MHQ validation studies in 15 languages. According to the evaluation process used here, the properties that achieved the best quality were responsiveness, hypothesis testing for construct validity and, to a lesser extent, reliability. Responsiveness was rated *sufficient* in three of the five versions analyzed, methodological quality *very good* in four, and quality of evidence *moderate* in three, with the best results. The property that achieved the lowest quality was measurement error, with the three studies that reported it rating it *doubtful* for methodological quality, *insufficient* for good measurement property, and *moderate* to *low* for quality of evidence.

Although moderate or high scores were found, cross-cultural validity was not developed in any version. Of the six properties that could be analyzed, structural validity, measurement error and responsiveness were rarely reported. Only the Thai version by Atthakomol et al. [26] developed all six psychometric properties and achieved high quality scores.

In terms of psychometric properties, structural validity was only analyzed in three versions [26–28] which is consistent with other reviews of hand PROMs, such as the PRWE [44] with three versions that developed this property, or the review of the Jebsen-Taylor Hand Function Test [46], where it was not developed it in any of the versions. It would therefore be interesting to include this property in future versions of the MHQ, as it directly influences internal consistency [17,20].

Internal consistency was considered in all versions, although Cronbach's alpha values were not higher than 0.7 in any reported data. In contrast, in the original version [5], in revisions of versions of the PRWE [44] or neck-specific questionnaires [45] Cronbach's alpha was higher 0.7. High values indicate the ability to detect small expected changes both clinically, as after analgesic treatment in patients with osteoarthritis, and over time, as in long processes such as Dupuytren's disease [18].

However, the quality of evidence could not be assessed, because the majority of versions lacked structural validity. Similar findings were reported by Shafiee et al. [46] in their review of the Patient-Rated Tennis Elbow Evaluation (PRTEE), which found indeterminate ratings and low quality of evidence for structural validity and internal consistency.

Test-retest reliability was assessed on ICC, which was mostly above 0.7 (acceptable) for all versions, except for some domains [28,32,33]. High ICC values are associated with high assessment and follow-up scores on the original MHQ [47]. The test-retest time interval was taken into account, varying from one week in the original version [5] to three days [15] or two weeks in other versions [27,32–34,48]. The recommendation of two weeks by Mokkink et al. [20] was chosen, which influenced the assessment of methodological quality as doubtful.

Particularly for the MHQ, good development of validity and reliability is important, as these are properties that influence the instrument's characteristics as a discriminative and diagnostic PROM [47].

Table 3
Methodological quality and ratings of measurement properties of the selected cross-cultural versions of MHQ.

PROM	Version	Structural validity	Internal consistency	Reliability	Measurement error	Hypothesis testing ¶	Responsiveness
Miranda et al. 2008 [28]	Colombian Spanish						
Criteria for good measurement property* (Rating)		?	7+	3+/3-	NR	"+" SF-36 Physical/SF-36 Pain "-" SF-36 Role physical	?
Methodological quality. Risk of bias		Adequate	Very good	Doubtful	NR	Adequate SF-36	Doubtful
van der Giesen et al. 2008 [42]	Dutch						
Criteria for good measurement property* (Rating)		NR	NR	NR	NR	NR	5+/2-
Methodological quality. Risk of bias		NR	NR	NR	NR	NR	Very good
Öksüz et al. 2011 [31]	Turkish						
Criteria for good measurement property* (Rating)		NR	?	11+	NR	"+" DASH/DASH-W/DASH S-M/VAS/Grip	NR
Methodological quality. Risk of bias		NR	Doubtful	Inadequate	NR	Adequate DASH/DASH-W/DASH S-M/VAS/Grip	NR
Roh et al. 2011 [38]	Korean						
Wi et al 2013 [41]	Korean						
Criteria for good measurement property* (Rating)		NR	?	12+	NR	"+" DASH/DASH-W/DASH S-M/VAS/FS/Grip. "-" ROM	NR
		NR	NR	NR	NR	NR	7+
Methodological quality. Risk of bias		NR	Doubtful	Doubtful	NR	Adequate DASH. Inadequate VAS/FS/ROM/Grip	NR
		NR	NR	NR	NR	NR	Very Good
Meireles et al. 2014 [32]	Brazilian Portuguese						
Criteria for good measurement property* (Rating)		NR	?	27+/1-	NR	"+" HAQ/DASH/DASH-W/Cochin/VAS	NR
Methodological quality. Risk of bias		NR	Doubtful	Doubtful	NR	Adequate HAQ/ Cochin/VAS Very good DASH	NR
Marks et al. 2014 [35]	German					.	
Criteria for good measurement property* (Rating)		NR	?	7+	5+/2-	"+" DASH/SF-12 P/K-Pinch "-" SF-12 Mental	7+
Methodological quality. Risk of bias		NR	Doubtful	Doubtful	Doubtful	Very good DASH/SF12 P/K-Pinch. Doubtful SF-12 Mental	Very good
Ebrahimzadeh et al. 2015 [15]	Persian						
Criteria for good measurement property* (Rating)		NR	?	7+	NR	"+" DASH/DASH-W/DASH S-M. "-" DASH/VAS	NR
Methodological quality. Risk of bias		NR	Doubtful	Doubtful	NR	Adequate DASH/DASH-W Inadequate DASH S-M/VAS	NR
Ilhanli et al. 2015 [40]	Turkish						
Criteria for good measurement property* (Rating)		NR	?	7+	NR	? DASH/DASH-W	NR
Methodological quality. Risk of bias		NR	Doubtful	Doubtful	NR	Doubtful DASH/DASH-W	NR
Oda et al. 2016 [33]	Japanese						
Criteria for good measurement property* (Rating)		NR	?	12+/1-	NR	"+" DASH/SF-36	NR
Methodological quality. Risk of bias		NR	Doubtful	Doubtful	NR	Adequate DASH/SF-36 Inadequate SF-36 role emotional/Mental health	NR
Koziej et al. 2017 [36]	Polish						
Criteria for good measurement property* (Rating)		NR	?	15+	?	"+" DASH/DASH-W/VAS/Grip. "-" Pinch	NR
Methodological quality. Risk of bias		NR	Doubtful	Doubtful	Doubtful	Adequate DASH/DASH-W/VAS/Grip/Pinch	NR
Dhippayom et al. 2018 [34]	Thai						
Criteria for good measurement property* (Rating)		NR	?	7+	NR	?	NR
Methodological quality. Risk of bias		NR	Doubtful	Doubtful	NR	Inadequate	NR
Busuioc et al. 2018 [30]	Canadian French						
Criteria for good measurement property* (Rating)		NR	?	15+	NR	"+" DASH/SF-12 Physical/VAS/DASH S-M/DASH-W. "-" SF-12 Mental	NR
Methodological quality. Risk of bias		NR	Doubtful	Inadequate	NR	Adequate DASH/SF-12 Physical/VAS/DASH S-M/DASH-W. SF-12 Mental	NR
Hulkkonen et al. 2018 [27]	Finnish						
Uimonen et al. 2020 [39]	Finnish						
Criteria for good measurement property* (Rating)		NR	12 +	11+/1-	NR	"+" DASH/DASH-W/EQ-5D/EQ-VAS/VAS/Grip EQ-5D/EQ-VAS/VAS-Pain "-" DASH S/M	NR
		NR	NR	NR	NR	"+" EQ-5D Index/Self-care/usual activities/pain "-" EQ-5D mobility	NR

Table 3 (*Continued*)

PROM	Version	Structural validity	Internal consistency	Reliability	Measurement error	Hypothesis testing ¶	Responsiveness
Methodological quality. Risk of bias		Inadequate	Doubtful	Doubtful	NR	Adequate DASH/EQ-5D/VAS/Grip	NR
		NR	NR	NR	NR	Very good EQ-5D	NR
Atthakomol et al. 2020 [26]	Thai						
Criteria for good measurement property* (Rating)		+	12 +	7+	1-	"+" DASH/EQ-5D/Psychological DASH/EQ-5D anxiety "-" EQ-5D Mobility	6+
Methodological quality. Risk of bias		Doubtful	Very good	Doubtful	Doubtful	Adequate DASH/EQ-5D	Very Good
Passiatore et al. 2021 [54]	Italian						
Criteria for good measurement property* (Rating)		NR	?	11+	NR	"+" DASH/DASH-W/VAS/Grip	NR
Methodological quality. Risk of bias		NR	Doubtful	Doubtful	NR	Adequate DASH/DASH-W/VAS/Grip	NR
Lye et al. 2021 [48]	Malay						
Criteria for good measurement property* (Rating)		NR	?	11+	NR	"+" DASH	NR
Methodological quality. Risk of bias		NR	Doubtful	Doubtful	NR	Adequate DASH	NR
Blomstrand et al. 2021 [29]	Swedish						
Criteria for good measurement property* (Rating)		NR	?	7+	NR	"+" PRWE-Pain/PRWE-Activity/VAS	NR
Methodological quality. Risk of bias		NR	Doubtful	Inadequate	NR	Adequate PRWE/VAS	NR
Kumar et al. 2022 [37]	Malaysian						
Criteria for good measurement property* (Rating)		NR	?	7+	NR	?	NR
Methodological quality. Risk of bias		NR	Doubtful	Doubtful	NR	Inadequate	NR

NR: not reported; NE: not evaluable; – = insufficient; + = sufficient;? = indeterminate; SF-36 = Short Form -36; SF-12 = Short Form -12; SF-12 P = Short Form -12 Physical. DASH = Disability of Arm, Shoulder and Hand; DASH-W = Disability of Arm, Shoulder and Hand-Work; DASH S-M = Disability of Arm, Shoulder and Hand Sport-Music; VAS = Visual Analogue Scale; FS = Function Scale; ROM = Range of Motion; HAQ= Health Assessment Questionnaire; K-Pinch = Key Pinch; PRWE = Patient-Rated Wrist Evaluation; EQ-5D = EuroQol -5Dimensions.

* The results of each study evaluating a measurement property of a PROM were rated against the updated criteria for good measurement properties (rating): – = insufficient; + = sufficient;? = indeterminate.

¶ The hypotheses tested were either confirmed "+" or rejected "–" for each of the studies.

Measurement error is important in the assessment of function and change and in decision making [44] [23]. This was not included in the original version, but was analyzed in three versions [26,35,36] with SEM values considered low [49]: from 3.9 in patients with osteoarthritis [35] to 10.7 in patients with conservatively treated hand conditions [36]. Similar results are seen in other hand PROMs such as the PRWE, which ranges from 5.22 to 8.12 in distal radius fractures and resection arthroplasty across various upper-extremity conditions [44].

All versions included construct validity, using Spearman *rho* and Pearson *r* correlation coefficients. Moderate to high correlations were, as expected, found between instruments measuring similar constructs, such as the DASH and its scales and the PRWE. Additionally, moderate to weak correlations were found with the SF-12, SF-36 and EQ-5D. Particularly weak correlations were found with measures more related to clinical records, such as grip, key pinch or pain on VAS. These findings are consistent with other reviews of functional hand impairment [18,44,47] and suggest that performance for activities of daily living does not depend solely on physical factors such as strength or range of motion [18,47]. Although a-priori hypotheses were presented for each study, there were 8 versions [15,29,30,32–34,37,50] in which no hypotheses were presented, indicating a high risk of bias [47,51].

For instruments such as the MHQ, which are more suitable for diagnostic purposes, responsiveness is not considered a key property [47], although it was evaluated in five versions. Effect sizes and SRMs ranged from 0.6 [41] to 1.7 [35], which are considered moderate to high. Kotsis et al. [6] found MHQ responsiveness scores ranging from 0.5 to 1.1. Similarly, methodological quality was assessed in the review by Arcidiacone et al. [19], and rated high. These findings suggest that the MHQ is responsive to changes brought about by clinical interventions in patients with hand conditions, allowing differentiation between those who experience improvement and those who do not.

4.1. Strengths and limitations

This was the first study to analyze the cross-cultural versions of the MHQ, one of the most important measurement tools in hand surgery and hand rehabilitation. Other reviews have been published on versions of the MHQ in clinical settings [18,19], but not specifically on cross-cultural validation scenarios with their own characteristics and procedures that ensure their applicability in different countries and languages. In this respect, the MHQ was examined by Shauver et al. [18] and Arcidiacone et al. [19], who used different methodologies or focused solely on methodological quality in their analyses.

The present review provides a transparent synthesis, with information from all versions in a single document. Although there may be some subjective criteria in the analysis, these were counterbalanced by assessment by two independent reviewers. However, the results of the studies presented in this review had their own limitations, such as lack of analysis of structural validity in most versions. Other aspects of applicability, such as "worst score counts" principle or the absence of "adequate" internal consistency, limited the precision of rating in terms of methodological quality. Although content validity assessment is recommended [52], it requires a separate detailed study that is beyond the scope of this research.

5. Conclusions

The current state of methodological quality, evidence and good measurement properties of the various cross-cultural versions of the MHQ is moderate or high for responsiveness and hypothesis

Table 4
Synthesized evidence, overall rating for good measurement properties and quality of evidence.

PROM	Measurement property	Summarized result	Overall Rating*	Quality of evidence §
Miranda et al. 2008/ Colombian Spanish version [29]	Structural validity	CTT: CFA: CFIor TLI/RMSEA or SRMR not calculated	?	
	Internal consistency	Cronbach's α (0.78–0.96)	+	High
	Reliability	ICC<0.70 D1, D4, D3	±	
	Hypotheses testing for construct validity	2 out of 3 hypotheses were confirmed	–	Moderate
	Responsiveness	Not ES or SRM or Normaṅs coefficient or t-statistic applied	?	
van der Giesen et al.2008/ Dutch version [43]	Responsiveness	Sample size (n = 28). The ES or SRM is smaller (<0.5) in 71% of the results.3 hypotheses were not confirmed	–	Low
Öksüz et al. 2011/Turkish version [32]	Internal consistency	Cronbach's α (0.76–0.96). Unestablished structural validity	?	
	Reliability	ICC>0.7 (0.79–0.96)	+	Moderate
	Hypotheses testing for construct validity	Spearmaṅs >0.3 similar construct and <0.3 different construct 5 out of 5 hypotheses were confirmed	+	High
Roh et al. 2011/Korean version [39]	Internal consistency	Cronbach's α (0.79–0.97). Unestablished structural validity.	?	
	Reliability	ICC>0.70 (0.88–0.96)	+	Moderate
	Hypotheses testing for construct validity	Spearmaṅs >0.3 similar construct and <0.3 different construct. 4 out of 5 hypotheses were confirmed	+	Moderate
Wi et al. 2013/Korean version [42]	Responsiveness	Hypotheses testing: before and after intervention/ comparison between subgroups Hypotheses ES or SRM > 0.5 confirmed	+	Moderate
Meireles et al.2014/ Brazilian-Portuguese version [33]	Internal consistency	Cronbach's α (0.969–0.864). Unestablished structural validity.	?	
	Reliability	ICC>0.7; IRE1 (0.84–0.98); IRE2 (0.61–0.93)	+	Moderate
	Hypotheses testing for construct validity	Spearmaṅs >0.3 similar construct and <0.3 different construct. 5 out of 5 hypotheses were confirmed. Sample size <50	+	Low
Marks et al. 2014/German version [36]	Internal consistency	Cronbach's α (0.77–0.89). Unestablished structural validity	?	
	Reliability	ICC>0.7 (0.85–0.95)	+	Moderate
	Measurement error	SDC < MIC 5 +out of 7 results. < 75%	–	Low
	Hypotheses testing for construct validity	Pearson>0.3 similar construct and <0.3 different construct. 3 out of 4 hypotheses were confirmed	+	High
	Responsiveness	Results in accordance with the hypothesis. ES ≥ 0.5." ES (1.9–0.5) SRM (1.8 a 0.4)	+	Moderate
Ebrahimzadeh et al. 2015/ Persian version [16]	Internal consistency	Cronbach's α (0.63–0.96). Unestablished structural validity	?	
	Reliability	ICC>0.7 (0.73–0.86)	+	Moderate
	Hypotheses testing for construct validity	3 out of 4 hypotheses were confirmed	+	Moderate
Ilhanli et al. 2015/Turkish version [41]	Internal consistency	Cronbach's α (0.63–0.96). Unestablished structural validity	?	
	Reliability	ICC>0.7 (0.98–0.99)	+	Moderate
	Hypotheses testing for construct validity	No correlation results with another scale were provided	?	
Oda et al. 2016/ Jappanese version [34]	Internal consistency	Cronbach's α > 0.70 Unestablished structural validity	?	
	Reliability	ICC>0.70 (0.683–-0.99) for 12 items except for D5 left 0.683. Overall>75%	+	Moderate
	Hypotheses testing for construct validity	7 out of 9 hypotheses were confirmed	+	Moderate
Koziej et al. 2017/Polish version [37]	Internal consistency	Cronbach's α > 0.70(0.78–0.96) Unestablished structural validity.	?	
	Reliability	ICC>0.7 (0.82–0.97)	+	Moderate
	Measurement error	MIC was not defined	?	
	Hypotheses testing for construct validity	5 out of 5 hypotheses were confirmed	+	Moderate
Dhippa-yom et al. 2018/ Thai version [35]	Internal consistency	Cronbach's α > 0.70; D5 Cronbach's α < 0.70 Unestablished structural validity	?	
	Reliability	ICC >0.70 (0.88–0.96). Sample size <50	+	Low
	Hypotheses testing for construct validity	Hypothesis not applicable to any comparator instrument or comparison between subgroups	?	
Busuioc et al. 2018/ Canadian-French version [31]	Internal consistency	Cronbach's α > 0.70. Unestablished structural validity	?	
	Reliability	ICC>0.7 (0.81–0.95)	+	Moderate
	Hypotheses testing for construct validity	5 out of 6 hypotheses were confirmed	+	Moderate
Hulkkonen et al. 2018/ Finnish version [28]	Structural validity	CTT for CFA: CFI = 0.84 and TLI = 0.83	–	Low
	Internal consistency	Cronbach's α > 0.70 (0.79–0.97)	+	Moderate
	Reliability	ICC>0.70 for 11 out of 12 results	+	Moderate
	Hypotheses testing for construct validity	6 out of 8 hypotheses were confirmed	+	Moderate

Table 4 (*Continued*)

PROM	Measurement property	Summarized result	Overall Rating*	Quality of evidence §
Uimonen et al. 2020/ Finnish version [40]	Hypotheses testing for construct validity	7 out of 8 hypotheses were confirmed	+	High
Atthakomol et al. 2020/Thai version [27]	Structural validity	CFI = 0.985; TLI = 0.983; RMSEA of 0.080 (90%C.I., 0.073–0.087), SRMR of 0.067/unidimensional CFI = 0.995, TLI = 0.994, RMSEA = 0.047 (90%C.I.0.039–0.055), SRMR = 0.041, /6 factors	+	Moderate
	Internal consistency	Cronbach's α > 0.70 (0.79−0.98)	+	High
	Reliability	ICC>0.70 (0.83−0.95)	+	Moderate
	Measurement error	SDC = 14.4; MIC = 5.2 SDC > MIC	–	Moderate
	Hypotheses testing for construct validity	9 out of 10 hypotheses were confirmed	+	Moderate
	Responsiveness	ES = 0.69; SRM = 0.78. Results in line with the hypotheses in comparison between subgroups and before and after intervention	+	Moderate
Passiatore et al. 2021/ Italian version 58	Internal consistency	Cronbach's α > 0.70 (0.91−0.99). Unestablished structural validity	?	
	Reliability	ICC>0.70 (0.87−0.98)	+	Moderate
	Hypotheses testing for construct validity	3 out of 4 hypotheses were confirmed	+	Moderate
Lye et al. 2021/ Malay version [50]	Internal consistency	Cronbach's α > 0.70 (0.79−0.98). Unestablished structural validity.	?	
	Reliability	ICC>0.70 (0.925−0.984)	+	Moderate
	Hypotheses testing for construct validity	Results in line with 1 hypothesis	+	Moderate
Blomstrad et al. 2021/ Swedish version [30]	Internal consistency	Cronbach's α (T1/T2) (0.77−0.96). Unestablished structural validity	?	
	Reliability	ICC>0.70 (0.77−0.92)	+	Moderate
	Hypotheses testing for construct validity	3 out of 3 hypotheses were confirmed	+	Moderate
Kumar et al. 2022/Malayan version [38]	Internal consistency	Cronbach's α (T1/T2) (0.96−0.77). Unestablished structural validity.	?	
	Reliability	ICC>0.70 (0.960−0.984)	+	Moderate
	Hypotheses testing for construct validity	The hypotheses were not applicable to other comparator instrument or comparison between subgroups	?	

– = insufficient; + = sufficient;? = indeterminate; ± inconsistent; CTT = Classical Test Theory; CFA = Confirmatory Factor Analysis; CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; RMSEA = Root Mean Square Error of Approximation; SRMR = Standardized Root Mean Square Residual; ICC = Intraclass Correlation Coefficient; D = Dimension; ES = Effect size; SRM = Standardized Response Mean; IRE1 = Inter-rater evaluation; IRE2 = Intra-rater evaluation; SDC = Smallest Detectable Change; MIC = Minimum Important Change.

* The outcomes from the studies pertaining to a specific measurement property of a PROM were summarized qualitatively and then rated against the updated criteria for good measurement properties: – = insufficient; + = sufficient; ± = inconsistent;? = indeterminate.

§ The evaluation of evidence quality was conducted using a GRADE approach.

testing for construct validity. Precision is limited by the lack of development of other properties such as measurement error or the formulation of a-priori hypotheses. Failure to include structural validity significantly limits the internal consistency of a large proportion of the MHQ versions, affecting the detection of small clinical changes and their discriminative capacity. We therefore suggest that future versions develop these measurement properties, with larger sample sizes and longer test-retest intervals.

Author contributions

MVMF: Design, methodology, writing - original draft. ISH: Software and validation. JMC: Formal analysis. Sarabia Cobo C: desing, writing - review and editing resources, supervision.

Funding

This study did not receive any grant from funding agencies in the public, commercial, or non-profit sectors.

Human rights

The study was conducted in accordance with the Declaration of Helsinki of the World Medical Association, revised in 2013, for experiments involving humans.

Informed consent and patient details

The authors confirm that they obtained written informed consent from the patients and/or volunteers featured in the article. Furthermore, they ensure that this report does not contain any personal information that could potentially reveal patients' identity.

Conflicts of interest

The authors declare no conflict of interest; the study has not been presented at any meeting.

Ethics considerations

To conduct this systematic review, we followed the methodological criteria outlined in the current consensus-based standards for the selection of health measurement instruments COSMIN for systematic reviews. In addition, the general recommendations of the PRISMA statement were followed. This study has been registered in the PROSPERO database for systematic reviews (CRD no.: 42022303561).

Acknowledgment

Special thanks to Karen Valish and Dr. Jaime Sánchez-Ruas for their invaluable support in the early stages of this process.

References

[1] Cheung K, Hatchell A, Thoma A. Approach to traumatic hand injuries for primary care physicians. Can Fam Physician 2013;59:614–8.

[2] Weldring T, Smith SMS. Patient-Reported Outcomes (PROs) and Patient-Reported Outcome Measures (PROMs). Health Serv Insights 2013;6:61–8. http://dx.doi.org/10.4137/HSI.S11093.

[3] Veehof MM, Sleegers EJA, Van Veldhoven NHMJ, Schuurman AH, Van Meeteren NLU. Psychometric qualities of the dutch language version of the disabilities of the arm, shoulder, and hand questionnaire (DASH-DLV). J Hand Ther 2002;15:347–54. http://dx.doi.org/10.1016/s0894-1130(02)80006-0.

[4] Roh YH. Clinical evaluation of upper limb function: patient's impairment, disability and health-related quality of life. J Exerc Rehabil 2013;9:400–5. http://dx.doi.org/10.12965/jer.130060.

[5] Chung KC, Pillsbury MS, Walters MR, Hayward RA. Reliability and validity testing of the Michigan Hand Outcomes Questionnaire. J Hand Surg Am 1998;23:575–87. http://dx.doi.org/10.1016/S0363-5023(98)80042-7.

[6] Kotsis SV, Lau FH, Chung KC. Responsiveness of the Michigan Hand Outcomes Questionnaire and physical measurements in outcome studies of distal radius fracture treatment. J Hand Surg Am 2007;32:84–90. http://dx.doi.org/10.1016/j.jhsa.2006.10.003.

[7] Sambandam SN, Priyanka P, Gul A, Ilango B. Critical analysis of outcome measures used in the assessment of carpal tunnel syndrome. Int Orthop 2008;32:497–504. http://dx.doi.org/10.1007/s00264-007-0344-7.

[8] Chatterjee JS, Price PE. Comparative responsiveness of the Michigan hand outcomes questionnaire and the carpal tunnel questionnaire after carpal tunnel release. J Hand Surg Am 2009;34:273–80. http://dx.doi.org/10.1016/j.jhsa.2008.10.021.

[9] McMillan CR, Binhammer PA. Which outcome measure is the best? Evaluating responsiveness of the Disabilities of the Arm, Shoulder, and Hand questionnaire, the Michigan Hand questionnaire and the Patient-Specific Functional Scale following Hand and wrist surgery. Hand 2009;4:311–8. http://dx.doi.org/10.1007/s11552-009-9167-x.

[10] Wehrli M, Hensler S, Schindele S, Herren DB, Marks M. Measurement properties of the brief Michigan hand outcomes questionnaire in patients with dupuytren contracture. J Hand Surg Am 2016;41:896–902. http://dx.doi.org/10.1016/j.jhsa.2016.06.009.

[11] Hudak PL, Amadio PC, Bombardier C. Development of an upper extremity outcome measure: the DASH (Disabilities of The Arm, Shoulder, and Hand). Am J Ind Med 1996;29:602–8. http://dx.doi.org/10.1093/oso/9780198749394.003.0008.

[12] Horng Y-S, Lin M-C, Feng C-T, Huang C-H, Wu H-C, Wang J-D. Responsiveness of the Michigan hand outcomes questionnaire and the disabilities of the arm, shoulder, and hand questionnaire in patients with hand injury. J Hand Surg Am 2010;35:430–6. http://dx.doi.org/10.1016/j.jhsa.2009.11.016.

[13] Black RM. Cultural considerations of hand use. J Hand Ther 2011;24:104–11. http://dx.doi.org/10.1016/j.jht.2010.09.067.

[14] Beaton D, Bombardier C, Guillemin F, Ferraz M. Guidelines for the process of cross-cultural adaptation of self-report measures. Spine (Phila Pa 1976;2000(25):3186–91. http://dx.doi.org/10.1097/00007632-200012150-00014.

[15] Ebrahimzadeh MH, Birjandinejad A, Kachooei AR. Cross-cultural adaptation, validation, and reliability of the Michigan Hand Outcomes questionnaire among Persian population. Hand Surg 2015;20:25–31. http://dx.doi.org/10.1142/S0218810415500033.

[16] Michigan Center for Hand Outcomes and Innovation Research. MHQ Traslations. Michigan Center for Hand Outcomes and Innovation Research n.d. www.mchoirresearch.wixsite.com/themhq/traslations (accessed September 20, 2022).

[17] Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. Qual Life Res 2018;27:1147–57. http://dx.doi.org/10.1007/s11136-018-1798-3.

[18] Shauver MJ, Chung KC. The Michigan hand outcomes questionnaire after 15 years of field trial. Plast Reconstr Surg 2013;131:779–87. http://dx.doi.org/10.1097/PRS.0b013e3182865d83.

[19] Arcidiacone S, Panuccio F, Tusoni F, Galeoto G. A systematic review of the measurement properties of the Michigan Hand Outcomes Questionnaire (MHQ). Hand Surg Rehab 2022;41:542–51. http://dx.doi.org/10.1016/j.hansur.2022.08.005.

[20] Mokkink LB, de Vet HCW, Prinsen CAC, Patrick DL, Alonso J, Bouter LM, et al. COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. Qual Life Res 2017. http://dx.doi.org/10.1007/s11136-017-1765-4.

[21] Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:71. http://dx.doi.org/10.1136/bmj.n71.

[22] RAYYAN Teams. RAYYAN. 2023. https://www.rayyan.ai/ (accessed July 29, 2023).

[23] Mokkink LB, Boers M, Van Der Vleuten C, Patrick DL, Alonso J, Bouter LM, et al. COSMIN Risk of Bias tool to assess the quality of studies on reliability and measurement error of outcome measurement instrument user manual. Amsterdam 2021.

[24] Terwee CB, Mokkink LB, Knol DL, Ostelo RWJG, Bouter LM, De Vet HCW. 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. http://dx.doi.org/10.1007/s11136-011-9960-1.

[25] Mokkink LB, Terwee CB, Gibbons E, Stratford PW, Alonso J, Patrick DL, et al. Inter-rater agreement and reliability of the COSMIN (COnsensus-based Standards for the selection of health status Measurement Instruments) Checklist. Med Res Methodol 2010;10:70–82. http://dx.doi.org/10.1186/1471-2288-10-82.

[26] Atthakomol P, Manosroi W, Sanguanrungsirikul S, Punoppamas S, Benjachaya S, Tongprasert S, et al. A Thai version of the Michigan hand questionnaire (Thai MHQ): an investigation of the psychometric properties. Health Qual Life Outcomes 2020;18:313. http://dx.doi.org/10.1186/s12955-020-01548-0.

[27] Hulkkonen S, Repo JP, Häkkinen A, Karppinen J, Ryhänen J. Cross-cultural adaptation and validation of the Finnish version of the Michigan hand outcomes questionnaire. Scand J Surg 2020;109:159–65. http://dx.doi.org/10.1177/1457496918818981.

[28] Miranda D, Ramírez J, Rueda L, García J, Wolf G, LHL A. Validación del "Michigan Hand Outcomes Questionnaire" para población colombiana. Rev Colomb Reumatol 2008;15:271–90.

[29] Blomstrand J, Karlsson J, Fagevik Olsén M, Kjellby Wendt G. The Michigan Hand Outcomes Questionnaire (MHQ-Swe) in patients with distal radius fractures- cross-cultural adaptation to Swedish, validation and reliability. J Orthop Surg Res 2021;16:442. http://dx.doi.org/10.1186/s13018-021-02571-7.

[30] Busuioc SA, Karim M, Bourbonnais D, Efanov JI, Izadpanah A, Danino MA, et al. Cross-cultural adaptation, validity, reliability and clinical applicability of the michigan hand outcomes questionnaire, and its brief version, to Canadian French. J Hand Ther 2018;31:145–6. http://dx.doi.org/10.1016/j.jht.2017.11.008.

[31] Öksüz C, Akel BS, Oskay D, Leblebicioğlu G, Hayran KM. Cross-cultural adaptation, validation, and reliability process of the Michigan Hand Outcomes Questionnaire in a Turkish population. J Hand Surg Am 2011;36:486–92. http://dx.doi.org/10.1016/j.jhsa.2010.11.016.

[32] Meireles SM, Natour J, Lberton Batista DA, Lopes M, Skare TL. Cross-cultural adaptation and validation of the Michigan Hand Outcomes Questionnaire (MHQ) for Brazil: validation study. Sao Paulo Med J 2014;132:339–47. http://dx.doi.org/10.1590/1516-3180.2014.1326701.

[33] Oda T, Abe Y, Katsumi Y, Ohi H, Nakamura T, Inagaki K. Reliability and validity of the Japanese version of the Michigan hand outcomes questionnaire: a comparison with the DASH and SF-36 questionnaires. J Hand Surg Asian Pac Vol 2016;21:72–7. http://dx.doi.org/10.1142/S2424835516500119.

[34] Dhippayom JP, Trevittaya P, Cheng ASK. Cross-cultural adaptation, validity, and reliability of the patient-rated Michigan hand outcomes questionnaire for Thai patients. Occup Ther Int 2018;20188319875. http://dx.doi.org/10.1155/2018/8319875.

[35] Marks M, Audigé L, Herren DB, Schindele S, Nelissen RGHH, Vliet Vlieland TPM. Measurement properties of the German Michigan Hand Outcomes Questionnaire in patients with trapeziometacarpal osteoarthritis. Arthritis Care Res (Hoboken) 2014;66:245–52. http://dx.doi.org/10.1002/acr.22124.

[36] Koziej M, Trybus M, Mydłowska A, Sałapa K, Gniadek M, Banach M, et al. The Polish version of the Michigan Hand Outcomes Questionnaire: cross-cultural adaptation, reliability, construct validity, and measurement error. J Hand Surg Eur Vol 2018;43:199–208. http://dx.doi.org/10.1177/1753193417729579.

[37] Kumar A, Harun H, Hakim A, Ganapathy SSMA. Cross-cultural adaptation and validity of the patient rated Michigan hand outcome questionnaire in Bahasa Melayu for Malaysian patients. J Hand Surg Asian Pac Vol 2022;27:636–42. http://dx.doi.org/10.1142/S2424835522500606.

[38] Roh YH, Yang BK, Noh JH, Baek GH, Song CH, Gong HS. Cross-cultural adaptation and validation of the Korean version of the Michigan hand questionnaire. J Hand Surg Am 2011;36:1497–503. http://dx.doi.org/10.1016/j.jhsa.2011.06.006.

[39] Uimonen MM, Hulkkonen SM, Ryhänen J, Ponkilainen VT, Häkkinen AH, Karppinen J, et al. Assessment of construct validity of the Finnish versions of the disabilities of arm, shoulder and hand instrument and the Michigan hand outcomes questionnaire. J Hand Ther 2020;33:571–9. http://dx.doi.org/10.1016/j.jht.2019.03.008.

[40] Ilhanli I, Durmus D, Orekici G. Cultural adaptation of the Michigan hand outcomes questionnaire in patients with Carpal tunnel syndrome: a Turkish version study. Chronic Dis Transl Med 2015;1:42–7. http://dx.doi.org/10.1016/j.cdtm.2015.02.011.

[41] Wi SM, Gong HS, Bae KJ, Roh YH, Lee YH, Baek GH. Responsiveness of the Korean version of the Michigan hand outcomes questionnaire after carpal tunnel release. Clin Orthop Surg 2014;6:203–7. http://dx.doi.org/10.4055/cios.2014.6.2.203.

[42] van der Giesen FJ, Nelissen RG, Arendzen JH, de Jong Z, Wolterbeek R, Vliet Vlieland TP. Responsiveness of the Michigan Hand Outcomes Questionnaire-Dutch language version in patients with rheumatoid arthritis. Arch Phys Med Rehabil 2008;89:1121–6. http://dx.doi.org/10.1016/j.apmr.2007.10.033.

[43] Cohen J. Quantitative methods in psychology: a power primer. Psycholol Bull 1992;112:155–9. http://dx.doi.org/10.1037/0033-2909.112.1.155.

[44] Mehta SP, MacDermid JC, Richardson J, Macintyre NJ, Grewal R. A systematic review of the measurement properties of the patient-rated wrist evaluation. J Orthop Sports Phys Ther 2015;45:289–98. http://dx.doi.org/10.2519/jospt.2015.5236.

[45] Schellingerhout JM, Heymans MW, Verhagen AP, De Vet HC, Koes BW, Terwee CB. Measurement properties of translated versions of neck-specific questionnaires: a systematic review. BMC Med Res Methodol 2011;11:1–14. http://dx.doi.org/10.1186/1471-2288-11-87.

[46] Shafiee E, MacDermid JC, Walton D, Vincent JI, Grewal R. Psychometric properties and cross-cultural adaptation of the Patient-Rated Tennis Elbow Evaluation (PRTEE); a systematic review and meta-analysis. Disabil Rehabil 2022;44:5402–17. http://dx.doi.org/10.1080/09638288.2021.1938248.

[47] van de Ven-Stevens LA, Munneke M, Terwee CB, Spauwen PH, van der Linde H. Clinimetric properties of instruments to assess activities in patients with hand injury: a systematic review of the literature. Arch Phys Med Rehabil 2009;90:151–69. http://dx.doi.org/10.1016/j.apmr.2008.06.024.
[48] Lye JX, Kow RY, Ismail R, Khalid KA. The Malay version of the Michigan hand outcomes questionnaire: cross-cultural adaptation, validation and reliability testing. J Hand Surg Asian Pac Vol 2021;26:166–79. http://dx.doi.org/10.1142/S2424835521500156.
[49] Nunnally J, Bernstein I. Book Review: Psychometric Theory, 3rd ed, New York: McGraw-Hill; 1994.
[50] Efanov JI, Shine JJ, Darwich R, Besner Morin C, Arsenault J, Harris PG, et al. French translation and cross-cultural adaptation of the Michigan Hand Outcomes Questionnaire and the Brief Michigan Hand Outcomes Questionnaire. Hand Surg Rehabil 2018;37:86–90. http://dx.doi.org/10.1016/j.hansur.2017.12.003.
[51] Terwee CB, Bot SDM, de Boer MR, van der Windt DAWM, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007;60:34–42. http://dx.doi.org/10.1016/j.jclinepi.2006.03.012.
[52] Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol 2010;63:737–45. http://dx.doi.org/10.1016/j.jclinepi.2010.02.006.
[53] Ribeiro N, Schier AR, Cardoso A, Nardi AE. Traducao e adaptacao transcultural do Michigan Hand Outcomes Questionnaire. J Bras Psiquiatr 2011;60(2):99–110. http://dx.doi.org/10.1590/S0047-20852011000200005.
[54] Passiatore M, De Vitis R, Cilli V, Milano G, Saccomanno MF, Cotroneo C, et al. The Italian version of the Michigan Hand Outcomes Questionnaire (MHQ): translation, cross-cultural adaptation and validation. J Hand Surg Asian Pac Vol 2021;26(4):666–83. http://dx.doi.org/10.1142/S242483552150065X.