



HAL
open science

Measurement Properties and Minimal Important Change of the World Health Organization Disability Assessment Schedule 2.0 in Persons With Low Back Pain

Jessica J Wong, Astrid Desouza, Sheilah Hogg-Johnson, Wouter de Groot, Danielle Southerst, Melissa Belchos, Nadège Lemeunier, Stephanie Alexopoulos, Hamid Varmazyar, Silvano A Mior, et al.

► To cite this version:

Jessica J Wong, Astrid Desouza, Sheilah Hogg-Johnson, Wouter de Groot, Danielle Southerst, et al.. Measurement Properties and Minimal Important Change of the World Health Organization Disability Assessment Schedule 2.0 in Persons With Low Back Pain. Archives of Physical Medicine and Rehabilitation, 2023, 104 (2), pp.287-301. 10.1016/j.apmr.2022.06.005 . hal-04402274

HAL Id: hal-04402274

<https://ut3-toulouseinp.hal.science/hal-04402274v1>

Submitted on 18 Jan 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NoDerivatives 4.0 International License

REVIEW ARTICLE

Measurement Properties and Minimal Important Change of the World Health Organization Disability Assessment Schedule 2.0 in Persons With Low Back Pain: A Systematic Review



Jessica J. Wong, DC, MPH,^{a,b,c} Astrid DeSouza, MHSc,^{a,d} Sheilah Hogg-Johnson, PhD,^{a,b,e} Wouter De Groot, MD, PMR,^f Danielle Southerst, DC,^a Melissa Belchos, DC,^{a,c,d} Nadège Lemeunier, PhD,^{d,g} Stephanie Alexopoulos, DC,^{c,h} Hamid Varmazyar, MSc,ⁱ Silvano A. Mior, DC, PhD,^{a,d,e,h} Paula J. Stern, DC,^c Margareta C. Nordin, Dr. Med. Sci., PT,^{j,k} Anne Taylor-Vaisey, BA, MLS,^a Alarcos Cieza, PhD,^f Pierre Côté, DC, PhD^{a,b,d,h}

From the ^aInstitute for Disability and Rehabilitation Research, Ontario Tech University, Oshawa, Ontario, Canada; ^bDalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada; ^cDivision of Graduate Studies, Canadian Memorial Chiropractic College, Toronto, Ontario, Canada; ^dFaculty of Health Sciences, Ontario Tech University, Oshawa, Ontario, Canada; ^eDepartment of Research and Innovation, Canadian Memorial Chiropractic College, Toronto, Ontario, Canada; ^fRehabilitation Programme, Department of Noncommunicable Diseases, Sensory Functions, Disability, and Rehabilitation Unit, World Health Organization, Geneva, Switzerland; ^gEquipe constitutive du CERPOP, Université Toulouse III Paul Sabatier, Toulouse, France; ^hInstitute of Health Policy, Management, and Evaluation, University of Toronto, Toronto, Ontario, Canada; ⁱUndergraduate Education, Canadian Memorial Chiropractic College, Toronto, Ontario, Canada; ^jDepartment of Orthopedic Surgery, New York University, New York, NY; and ^kDepartment of Environmental Medicine, New York University, New York, NY.

Abstract

Objective: To determine the measurement properties and minimal important change (MIC) of the World Health Organization Disability Assessment Schedule 2.0 (WHODAS 2.0) short (12 questions) and full (36 questions) versions in persons with nonspecific low back pain (LBP).

Data Sources: MEDLINE, Embase, Cumulative Index to Nursing and Allied Health Literature, APA PsycInfo, and Cochrane Central Register of Controlled Trials (inception to May 2021).

Study Selection: Eligible studies assessed measurement properties or MIC of WHODAS 2.0 in persons with LBP.

Data Extraction: Paired reviewers screened articles, extracted data, and assessed risk of bias using Consensus-Based Standards for Selection of Health Measurement Instruments (COSMIN) and COSMIN–Outcome Measures in Rheumatology checklists.

Data Synthesis: We descriptively synthesized results stratified by measurement property and LBP duration (subacute: 6 weeks to 3 months; chronic: ≥ 3 months).

Results: We screened 297 citations and included 14 studies (reported in 15 articles). Methodological quality of studies was very good for internal consistency and varied between very good and doubtful for construct validity, doubtful for responsiveness, and adequate for all other properties assessed. Evidence suggests that WHODAS 2.0 full version has adequate content validity (2 studies); WHODAS 2.0 short and full versions have adequate structural validity (3 studies), but construct validity is indeterminate (9 studies). WHODAS 2.0 short and full versions have adequate

Funded by the Canadian Chiropractic Research Foundation (#2021-02). The funder was not involved in the design, data collection, analysis, interpretation, writing of the report, or the decision to submit the paper for publication. This research was undertaken, in part, thanks to funding from the Canada Research Chairs Program to Professor Pierre Côté, Canada Research Chair in Disability Prevention and Rehabilitation at Ontario Tech University.

Protocol Registration: OSF Registries (<https://osf.io/sxtgk>).

Disclosures: Dr Wong reports research grants from the Canadian Chiropractic Research Foundation (paid to institution); and the Canadian Institutes of Health Research (CIHR; paid to

institution) outside the submitted work. Dr Côté reports research grants from the Canadian Chiropractic Research Foundation (paid to institution); CIHR, College of Chiropractors of British Columbia, and the World Health Organization (paid to institution) outside the submitted work; payment for expert (court) testimony from the Canadian Chiropractic Protective Association and NCMIC; and support for attending meetings and/or travel from Eurospine, Sophiammet University, and World Federation of Chiropractic. Dr Nordin reports leadership role (unpaid) with World Spine Care, SPINE20, and Eurospine. The remaining authors declare that they have no conflicts of interest.

internal consistency (10 studies), and the full version has adequate test-retest and interrater reliability (3 studies) in persons with LBP. Minimal detectable change (MDC) was 10.45–13.99 of 100 for the full version and 8.6 of 48 for the short version in persons with LBP (4 studies). WHODAS 2.0 full version has no floor or ceiling effects, but the short version has potential floor effects in persons with chronic LBP (3 studies). One study estimated MIC for the full version as 4.87 of 100 or 9.74 of 100 (corresponding to 1- and 2-point change on 0- to 10-cm visual analog scale for pain, respectively), and 1 study estimated 3.09–4.68 of 48 for the short version.

Conclusions: In persons with LBP, WHODAS 2.0 full version has adequate content validity, structural validity, internal consistency, and reliability. WHODAS 2.0 short version has adequate structural validity and internal consistency. Construct validity of the short and full versions is indeterminate. Since MDC is estimated to be larger than MIC, users may consider both MIC and MDC thresholds to measure change in functioning for LBP.

Archives of Physical Medicine and Rehabilitation 2023;104:287–301

© 2022 by the American Congress of Rehabilitation Medicine. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Low back pain (LBP) is the leading cause of years lived with disability globally.¹ Global years lived with disability for LBP were 42.5 million in 1990 and increased by 53% to 64.9 million in 2017.² The global point prevalence of LBP was 7.8% in 2017, affecting 577 million people at any given time.² Overall, LBP has led to considerable disability and lost productivity across populations worldwide.^{1,3–6} Due to its high prevalence, LBP is also the main reason for unmet needs for rehabilitation globally.^{7,8}

Given the associated disability and needs for rehabilitation, it is important to have an instrument with adequate measurement properties to assess functioning in persons with LBP. The World Health Organization Disability Assessment Schedule 2.0 (WHODAS 2.0) is a self-reported generic questionnaire developed by the World Health Organization (WHO) to measure functioning across different health conditions, cultures, and settings.⁹ WHODAS 2.0 integrates an individual's level of functioning in major life domains and is directly linked to concepts of the International Classification of Functioning, Disability, and Health (ICF).⁹ Based on the ICF framework, functioning refers to all body functions and structures, activities and participation, and the interaction between a person's health condition(s) and that individual's contextual factors, including environmental and personal factors.⁹ However, the utility of WHODAS 2.0 in measuring functioning in persons with LBP remains unclear. Few systematic reviews have assessed the measurement properties of WHODAS 2.0 in persons with LBP or musculoskeletal conditions.^{10,11} Previous systematic reviews were limited by a narrow search strategy (eg, did not include subject headings for databases) or did not conduct risk of bias assessment of included studies.^{10,11} The most recent systematic review conducted a literature search that ended in January 2019,¹¹ and

relevant studies on the measurement properties of WHODAS 2.0 in persons with LBP have been published since early 2019.^{12–19}

Self-reported questionnaires to measure functioning are important tools for the management and rehabilitation of LBP. From a clinical perspective, health care providers can use a valid and reliable questionnaire to measure functioning to guide management, inform prognosis, and assess outcomes after rehabilitation interventions for LBP. At the population level, a valid and reliable questionnaire can be used in population-based surveys to measure functioning across different cultures and settings. However, most commonly used scales may not fully capture all constructs of functioning as per the ICF framework.²⁰ It is recommended that the measurement of outcomes in rehabilitation be based on the ICF.²¹ WHODAS 2.0 questionnaire is promising for providing a comprehensive view of functioning and the effect of disability in persons with LBP. WHODAS 2.0 covers the domains of cognition, mobility, self-care, getting along, life activities, and participation, as linked to these concepts in the ICF framework.⁹ WHODAS 2.0 questionnaire thus extends beyond the narrower focus of measuring functional limitations in LBP-specific questionnaires (eg, Oswestry Disability Index, Roland Morris Disability Questionnaire) and helps to compare the effect on functioning across different health conditions. Therefore, a systematic review is needed to comprehensively synthesize the evidence on the measurement properties of WHODAS 2.0 in persons with LBP.

We conducted a systematic review of the literature to determine the measurement properties (eg, validity, test-retest reliability including minimal detectable change [MDC], responsiveness), and minimal important change (MIC), of WHODAS 2.0 short (12 questions) and full (36 questions) versions in persons with nonspecific LBP. We conducted this systematic review alongside a scoping review (protocol: <https://doi.org/10.17605/OSF.IO/GQYVB>; to be published separately) to inform the development of a global indicator to measure effective coverage of LBP rehabilitation for the WHO to use in population-based surveys.

Methods

We registered our protocol with the Open Science Framework Registries on May 3, 2021 (<https://doi.org/10.17605/OSF.IO/SXTGK>). Guidance from Consensus-Based Standards for the Selection of Health Measurement Instruments (COSMIN) informed our design and conduct of the review; specifically, we incorporated COSMIN terminology for measurement properties and the COSMIN Risk of Bias checklist (with some modifications described below).²² We used the COSMIN–Outcome Measures in Rheumatology (OMERACT) checklist to appraise studies

List of abbreviations:

COSMIN	Consensus-Based Standards for the Selection of Health Measurement Instruments
ICC	intraclass correlation coefficient
ICF	International Classification of Functioning, Disability, and Health
LBP	low back pain
MDC	minimal detectable change
MIC	minimal important change
OMERACT	Outcome Measures in Rheumatology
SF-36	Short Form-36 Health Survey
WHO	World Health Organization
WHODAS 2.0	World Health Organization Disability Assessment Schedule 2.0

assessing the MIC of WHODAS 2.0.²³ We reported our systematic review according to the Preferred Reporting Items for Systematic Reviews and Meta-analyses statement.²⁴

Eligibility criteria

Population

We included studies of persons with nonspecific LBP. Nonspecific LBP is defined as pain between the costal margin and inferior gluteal folds with or without leg pain in the absence of underlying serious or major pathology (eg, fractures, dislocations, spinal cord injury, inflammatory arthritides, neoplasms, malignant neoplasms, any other serious pathology from trauma or deformities).²⁵ We focused on subacute (6-week to 3-month duration) and chronic (≥ 3 -month duration) LBP, given that the WHODAS 2.0 questionnaire has a recall period of 30 days. Study samples with musculoskeletal conditions or spinal pain were considered relevant if they included participants with nonspecific LBP (reporting results that included patients with LBP). We described these samples as having musculoskeletal conditions (including LBP) in the results section. We excluded studies targeting major neuromusculoskeletal or connective tissue disorders, autoimmune disorders, congenital disorders, major trauma (eg, with serious pathology), or deformities.

Instrument

We focused on WHODAS 2.0 questionnaire, which is a generic, self-reported assessment instrument developed by the WHO to provide a standardized method for measuring functioning across different cultures and settings.⁹ We included studies using the full (36 questions) (appendix 1) and short version (12 questions) (appendix 2) of the questionnaire. The measurement properties of WHODAS 2.0 have been examined in other conditions,^{10,11,26-29} and the short and full versions are correlated.^{30,31} We included studies using simple or complex scoring as well as summary and domain scores of WHODAS 2.0. Simple scoring involves adding up the scores from each WHODAS 2.0 item without recoding or collapsing response categories; thus, there is no weighting of individual items.⁹ Complex scoring involves item response theory-based scoring, which accounts for multiple levels of difficulty for each WHODAS 2.0 item, using an algorithm to determine the summary score by differentially weighting the items and levels of severity.⁹

Measurement properties

Measurement properties included validity, reliability, and responsiveness. Validity refers to the degree to which an outcome measure measures the construct(s) it purports to measure.²² Content validity is the degree to which the content of an outcome measure is an adequate reflection of the construct to be measured.²² Construct validity is the degree to which the scores of an outcome measure are consistent with hypotheses based on the assumption that the outcome measure validly measures the construct to be measured.²² Structural validity is the degree to which the scores of an outcome measure are an adequate reflection of the dimensionality of the construct to be measured.²² Reliability refers to the extent to which scores are measured without measurement error. For instance, for patients who have not changed, scores are the same for repeated measurements under several conditions (eg, over time [test-retest reliability] and when administered by different persons on the same

occasion [interrater reliability]).²² Responsiveness is the ability of an outcome measure to detect change over time in the construct to be measured when change has actually occurred.²² Interpretability, including MIC, is the degree to which one can assign qualitative meaning (clinical or commonly understood connotations) to a measure's quantitative scores or change in scores.²² MIC is defined as the smallest change in score that patients perceive as important.³² Interpretability includes evaluating the distribution of scores in the study population and floor and ceiling effects.²²

In addition to the above criteria, eligible studies met the following: (1) published in English, French, Chinese, German, Greek, Farsi, Swedish, or Italian languages (to increase feasibility) and (2) assessed measurement properties (eg, validity, reliability, responsiveness) or MIC of the WHODAS 2.0 short and full versions in persons with LBP. We excluded the following during citation screening: (1) guidelines, letters, editorials, commentaries, government reports, books and book chapters, conference proceedings, meeting abstracts, lectures and addresses, consensus development statements, guideline statements, and studies not reporting on methodology and (2) systematic reviews, literature reviews, and case studies.

Information sources and search strategy

We developed the search strategy in consultation with an experienced health sciences librarian. The search strategy was reviewed by a second librarian using the Peer Review of Electronic Search Strategies Checklist.³³ We searched MEDLINE, Embase, Cumulative Index to Nursing and Allied Health Literature, APA PsycInfo, and Cochrane Central Register of Controlled Trials. We used 2 searches focused on (1) key concepts of spinal pain and WHODAS 2.0 (searched databases from inception to April 22, 2021) (appendix 3) and (2) key concepts of musculoskeletal conditions and WHODAS 2.0 (searched databases from inception to May 10, 2021) (appendix 4). We incorporated 2 searches to form a comprehensive and sensitive search strategy that covered literature on spinal pain and musculoskeletal conditions more broadly. The search strategy was first developed in MEDLINE and adapted to other bibliographic databases. The searches combined subject headings (eg, Medical Subject Headings of the National Library of Medicine in MEDLINE) specific to databases with free text words related to WHODAS 2.0 and spinal pain or musculoskeletal conditions. We downloaded the search results into EndNote X9 and removed duplicate references electronically. We conducted a supplemental search of related systematic reviews to identify any additional relevant studies.^{10,11}

Study selection

We used a 2-phase screening process (titles/abstracts, full-text screening) to select eligible studies. We conducted training with screening a random sample of citations (50 titles/abstracts, 10 full texts) using predefined inclusion and exclusion criteria to achieve agreement of $\geq 80\%$ between reviewers before starting screening. In phase I screening, paired reviewers (J.J.W., A.D.) independently screened citation titles and abstracts to determine the eligibility of studies by categorizing studies as possibly relevant or irrelevant. Paired reviewers (J.J.W., A.D.) independently screened possibly relevant studies in full text during phase II screening to determine final eligibility and

document reasons for exclusion. Reviewers met to discuss disagreements and reach consensus on the eligibility of studies. A third reviewer (P.C.) was involved if consensus could not be reached. We contacted study authors for additional information as needed when screening, assessing risk of bias, and conducting data extraction of studies.

Methodological quality appraisal

We assessed the methodological quality of studies using the COSMIN checklist for studies assessing measurement properties and the COSMIN-OMERACT checklist for studies estimating MIC.^{23,34} Reviewers summarized judgments using the COSMIN and COSMIN-OMERACT checklists to assess the overall risk of bias for each study. All reviewers were trained in the use of these critical appraisal instruments, whereby training included (1) a review of the COSMIN and COSMIN-OMERACT checklists, notes, and manuals; (2) independent appraisals of an assigned article among reviewers; and (3) review and discussion of appraisals to reach consensus and clarify methodology if needed. Specifically, pairs of independent reviewers critically appraised eligible studies using the COSMIN and COSMIN-OMERACT checklists to inform their judgment on the internal validity of studies. For COSMIN, we assessed the quality of each study reporting on a measurement property with items including design requirements and preferred statistical methods, which were rated on a 4-point scale as “very good,” “adequate,” “doubtful,” or “inadequate.” We obtained a methodological quality score for each measurement property by taking the lowest rating across the items (ie, “the worst score counts” principle).³⁴ Pairs of reviewers discussed disagreements to reach consensus, and a third reviewer was involved if consensus could not be reached. When assessing content validity, we did not include a rating of the patient-reported outcome measure development and reviewer rating (modification from COSMIN methodology). The WHO developed the WHODAS 2.0 questionnaire in a range of populations, including the general population, populations with physical problems, mental or emotional problems, and problems related to alcohol and drug use.⁹ These studies were not restricted to LBP samples and were therefore outside the scope of our review.

Data extraction and synthesis

A reviewer extracted data from eligible studies to build evidence tables. A second reviewer verified all data items by checking the extracted data to minimize error. We extracted data from each study on author, year, study design, setting and participants, instrument, and measurement properties (eg, validity, reliability, responsiveness) and MIC where available. We calculated the MDC ($MDC=1.96 \times SEM \times \text{square root of } 2$; $SEM=SD \times \text{square root of } (1 - \text{intraclass correlation coefficient [ICC])$) using data from studies where available.³⁵ The percentage agreement and κ of agreement were computed for screening and critical appraisal. We descriptively synthesized results following the Synthesis Without Meta-analysis reporting guideline (appendix 5).³⁶ For the descriptive synthesis, we identified groups used in the synthesis, which were groupings based on measurement property, version of the questionnaire, and duration of LBP. We provided a description of the synthesized findings and certainty of the findings. We placed emphasis on measurement properties with methodological quality of very good or adequate and indicated which studies contributed to the synthesis.

Specifically, results were stratified by duration of LBP (eg, subacute: 6 weeks to 3 months vs chronic: ≥ 3 months), version of the questionnaire (short vs full version), and measurement property (eg, validity, reliability, MIC). Based on guidance from COSMIN,²² we interpreted reliability as “sufficient” when ICC or weighted $\kappa \geq 0.70$, “indeterminate” when ICC or weighted κ was not reported, and “insufficient” when ICC or weighted $\kappa < 0.70$. For hypotheses testing for construct validity, we interpreted construct validity as “sufficient” when the result was in accordance with the hypothesis ($\geq 75\%$ of results of all studies in accordance with hypotheses), “indeterminate” when no hypothesis was defined, and “insufficient” when the result was not in accordance with the hypothesis provided by the investigators. As a modification of COSMIN methodology, we did not formulate our own hypotheses (eg, direction and relative magnitude of correlations with other measure[s]) because of the multidimensional nature of the WHODAS 2.0 questionnaire.

Results

Study selection

We screened 297 citations for inclusion (fig 1). We identified 14 studies reported in 15 articles as relevant, which were included in our descriptive synthesis.^{12-19,37-43} Pilot screening achieved 92% (titles/abstracts) and 90% (full texts) agreement based on independent results. Interrater agreement for phase I and II screening based on independent results was 88% agreement; $\kappa=0.62$; 95% CI, 0.44-0.80 for search 1; and 92% agreement; $\kappa=0.62$; 95% CI, 0.47-0.78 for search 2. Independent critical appraisal before discussion to reach consensus had 80% agreement. After risk of bias assessment, the methodological quality based on COSMIN was (1) internal consistency: 6 of 6 studies rated as very good, (2) structural validity: 3 of 3 studies rated as adequate, (3) cross-cultural validity/measurement invariance: 1 study rated as adequate, (4) reliability: 3 of 3 studies rated as adequate, (5) construct validity: ranged from doubtful to very good based on 7 studies, and (6) responsiveness: 2 of 2 studies rated as doubtful (table 1). We contacted authors of 4 studies^{13,38,39,43} for clarification on study methodology, of which 2 authors^{13,43} responded with additional information.

Study characteristics

Of the 14 included studies, 7 (50%) assessed WHODAS 2.0 full version and 7 (50%) assessed the short version (table 2). Mean age of study populations ranged from 45-71 years, and the proportion of women ranged from 42%-72%. Study populations were persons with chronic LBP (4 studies), LBP with mixed duration (1 study), and other (9 studies; eg, samples of musculoskeletal conditions that included participants with LBP). For studies with results specific to musculoskeletal conditions (including LBP), the proportion of participants reported with LBP ranged from 39%-62%. Most studies (79%) were conducted in the European region, 1 in the Western Pacific, and 1 in the African region. Almost all studies (13/14) were conducted in high-income countries.

A)

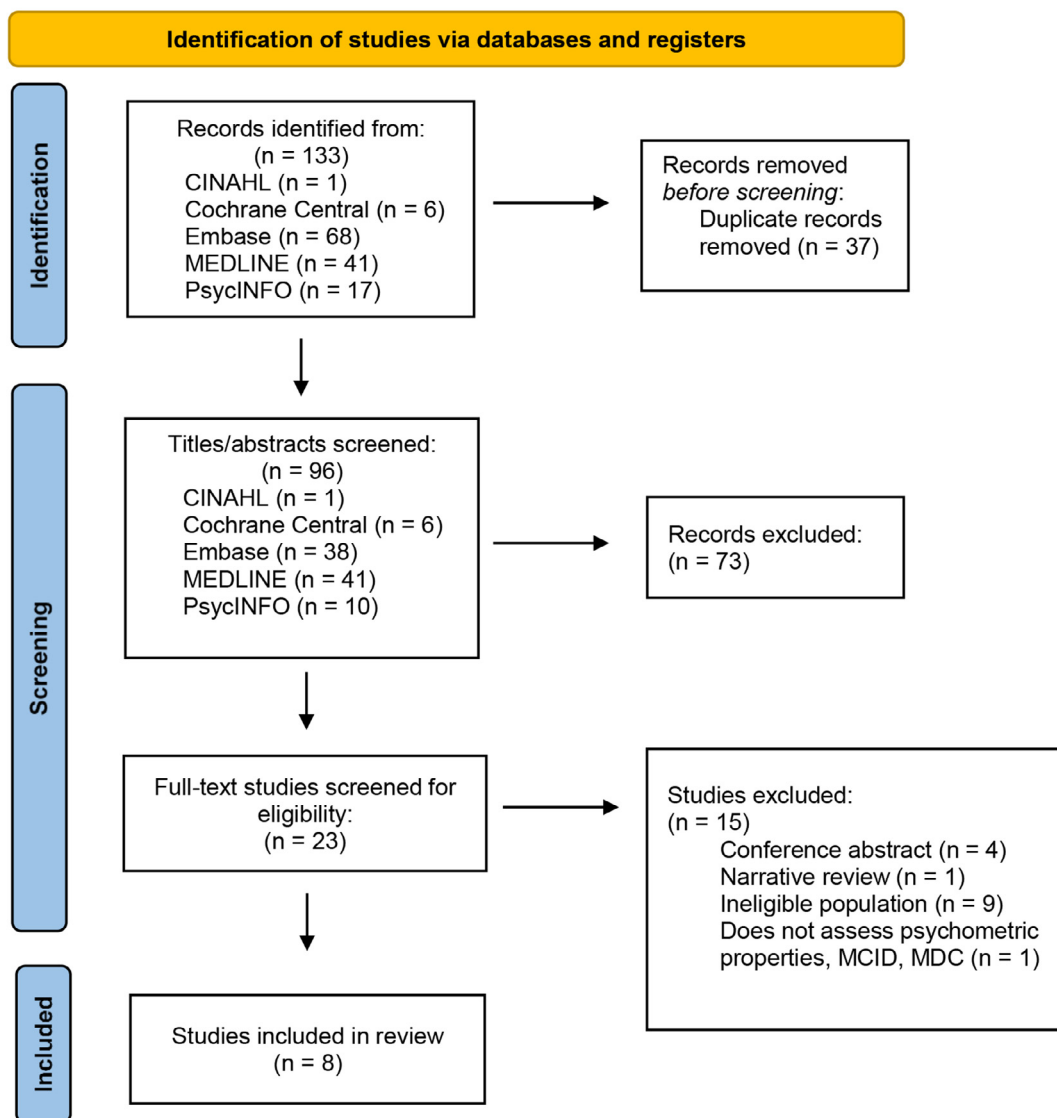


Fig 1 Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) flow diagram showing identification and selection of included studies for (A) search 1 (concepts of spinal pain and WHODAS 2.0) and (B) search 2 (concepts of musculoskeletal conditions and WHODAS 2.0). Abbreviations: CINAHL, Cumulative Index to Nursing and Allied Health Literature; MCID, minimum clinically important difference; MDC, minimal detectable change.

Validity

Content validity

Evidence from 2 studies suggests that WHODAS 2.0 full version has adequate content validity in persons with LBP^{15,40} (table 3). In 1 study of persons with acute or chronic LBP, 47 categories of the ICF were linked to WHODAS 2.0 full version, covering 30 of the 36 WHODAS 2.0 items.⁴⁰ Another study reported adequate cross-cultural adaptation of WHODAS 2.0 Igbo language full version among persons with chronic LBP in rural and urban Nigeria.¹⁵

Structural validity

Evidence from 3 studies (reported in 4 articles) suggests that the WHODAS 2.0 short^{41,42} and full^{15,39} versions have adequate structural validity in persons with LBP (see table 3).

Chronic LBP

In persons with chronic LBP, 1 study reported a 7-factor structure to WHODAS 2.0 Igbo language full version based on exploratory factor analysis (62.79% of items with factor loadings above 0.5; 66.67% of items loading on corresponding factor in the original measure).¹⁵

Other populations

In persons with musculoskeletal conditions of mixed duration (including LBP), the Kaiser-Mayer-Olkin's Measure (proportion

B)

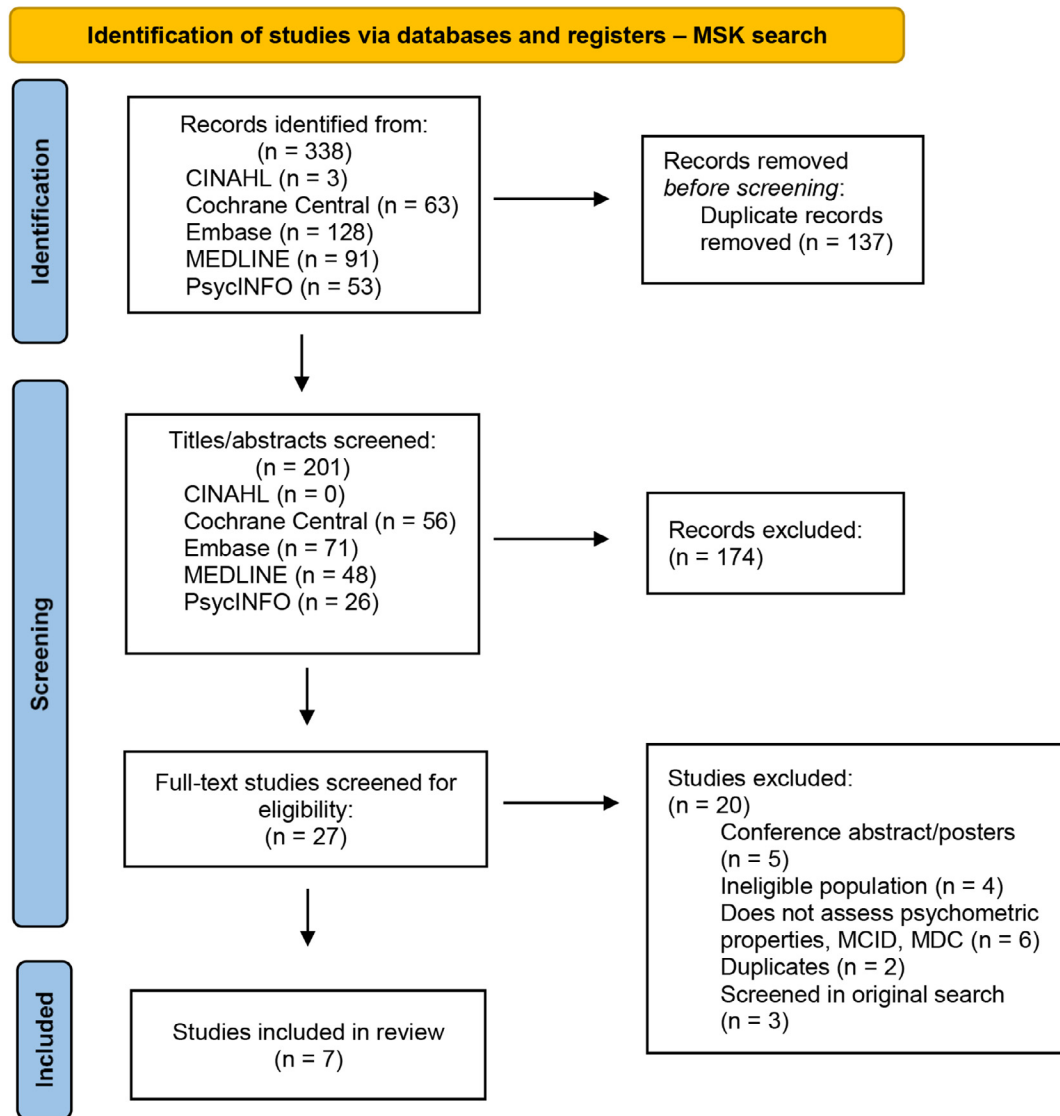


Fig 1 Continued

of variance in variables that may be caused by underlying factors) was 0.88 for WHODAS 2.0 German language full version.³⁹ In persons with chronic musculoskeletal conditions (including LBP), 1 study (reported in 2 articles) reported 2 retained factors to the Finnish language short version based on exploratory factor analysis (eigenvalues of 5.15 and 1.04) and a 2-factor structure based on confirmatory factor analysis (root mean square error of approximation 0.049, $\chi^2=1.99$, covariance between 2 common factors 0.79).^{41,42}

Construct validity for WHODAS 2.0 full version (36 questions)

Summary statement

Four of 5 studies stated a priori hypotheses related to the construct validity of WHODAS 2.0 full version (see table 3, appendix 6).^{13,15,38,39,43} The body of evidence from these 5 studies (3

supporting adequate construct validity, 1 indeterminate, 1 insufficient) suggests that construct validity of WHODAS 2.0 full version in persons with LBP is indeterminate.^{13,15,38,39,43}

Chronic LBP

In persons with chronic LBP, 1 study hypothesized that adults with a lower quality of life and lower mood would have higher levels of disability and that disability measured by Oswestry Disability Index and WHODAS 2.0 would correlate (strength of correlations not stated).¹³ A priori correlations were not hypothesized; thus, the construct validity is indeterminate. Moderate to strong correlations were reported between WHODAS 2.0 full version summary scores and Short Form-36 Health Survey (SF-36) domains ($r=-0.35$ to $r=-0.87$), Hospital Anxiety and Depression Scale domains ($r=0.46$ for anxiety, $r=0.49$ for depression), and Oswestry Disability Index ($r=0.87$).¹³ Another study hypothesized that WHODAS 2.0 summary scores would at least moderately correlate ($r\geq 0.3$) with pain and other measures of disability. This study reported

Table 1 Overall methodological quality for each measurement property of the WHODAS 2.0 short (12 questions) and full (36 questions) versions based on risk of bias assessment using the COSMIN Risk of Bias checklist

Author	WHODAS 2.0 Version	Measurement Property Assessed							
		Content Validity*	Structural Validity	Internal Consistency	Cross-Cultural Validity/ Measurement Invariance	Reliability	Criterion Validity	Construct Validity	Responsiveness
Chronic LBP									
Bärlund et al ¹²	Short							Very good	
Cwirlej-Sozanska et al ¹³	Full			Very good		Adequate		Doubtful	Doubtful
Goncalves Silva et al ³⁷	Short							Very good	
Igwesi-Chidobe et al ¹⁵	Full	Adequate	Adequate	Very good		Adequate		Very good	
Other									
Fujiwara et al ¹⁴	Short			Very good				Adequate	
Garin et al ³⁸	Full							Doubtful	
Katajapuu et al ¹⁶	Short			Very good					
Katajapuu et al ¹⁸	Short				Adequate				
Katajapuu et al ^{17,†}	Full								
Pösl et al ³⁹	Full		Adequate	Very good				Very good	Doubtful
Røe et al ⁴⁰	Full	Doubtful							
Saltychev et al ^{41,42}	Short		Adequate						
Silva et al ⁴³	Full			Very good		Adequate		Doubtful	
Tarvonen-Schröder et al ¹⁹	Short	Adequate						Very good	

Abbreviation: COSMIN, Consensus-based Standards for the Selection of Health Measurement Instruments.

* Did not include rating of patient-reported outcome measure development and reviewer rating for content validity (modification from COSMIN methodology).

† Study assessed floor and ceiling effects (informing interpretability only).

correlations of $r=0.54$ with Roland Morris Disability Questionnaire, $r=0.34$ with Back Performance Scale, and $r=0.56$ with 11-point Box Scale for pain.¹⁵ These results are in accordance with hypotheses, thus suggesting adequate construct validity.

Other populations

In persons with LBP, 1 study investigated known group validity and hypothesized differences in disability between groups with different levels of pain. No significant differences in WHODAS 2.0 full version summary scores were reported between persons with mild (numeric rating scale 0-3) vs severe (numeric rating scale 7-10) pain (effect size=0.45)³⁸; thus, the results are insufficient for construct validity. In samples of persons with musculoskeletal conditions of mixed duration (including LBP), 1 study hypothesized that subscales of WHODAS 2.0 and SF-36 that reflect the same dimensions would produce strong correlations (with a priori correlations stated).³⁹ Moderate correlations were reported between WHODAS 2.0 full version summary scores and SF-36 subscales ranging from $r=-0.43$ to $r=-0.62$ (highest correlation with SF-36 subscale Physical Functioning).³⁹ These results were in accordance with hypotheses and suggest adequate construct validity. In samples of persons with musculoskeletal conditions of mixed duration (including LBP), 1 study hypothesized that disability measured by WHODAS 2.0 increases with the number of pain points and pain intensity⁴³; the study assessed for a statistically significant correlation (magnitude of correlation not hypothesized a priori). A significant correlation was reported between WHODAS 2.0 full version summary scores and visual analog scale for pain intensity of $r=0.44$.⁴³ Mean \pm SD WHODAS 2.0 full version summary scores were 35.7 ± 13.3 among persons with widespread pain, 29.0 ± 14.1 in persons with pain in 2-3 sites,

and 24.1 ± 28.4 among persons with pain in only 1 location.⁴³ The results related to WHODAS 2.0 scores between different groups suggest adequate construct validity.

Construct validity for WHODAS 2.0 short version (12 questions)

Summary statement

Two of 4 studies reported a priori hypotheses to inform construct validity of WHODAS 2.0 short version (see table 3, appendix 6). The body of evidence from these 4 studies (2 supporting adequate construct validity, 2 indeterminate) suggests that the construct validity of WHODAS 2.0 short version in persons with LBP is indeterminate.^{12,14,19,37}

Other populations

In persons with chronic spinal pain, 1 study reported correlation between WHODAS 2.0 short version summary scores and WHO minimal generic data set scores was $r=0.48$ ($P<.05$); however, no hypothesis was provided.¹⁹ Another study aimed to investigate if WHODAS 2.0 short version could detect differences in disability between 2 different spinal conditions. Differences in WHODAS 2.0 short version summary scores were reported between persons with chronic spinal pain (mean, 22.0 ± 0.99) compared with persons with spinal cord injury (mean, 18.4 ± 0.99),¹⁹ in support of their expected findings. One study aimed to determine whether generic measures of disability can differentiate individuals with and without pain. In persons with musculoskeletal conditions of mixed duration (including LBP), there were significant differences in WHODAS 2.0 short version summary scores among persons with LBP (mean, 23.4 ± 8.7), persons with pain elsewhere (mean,

Table 2 Characteristics of included studies on the measurement properties and minimal important change of the WHODAS 2.0 short (12 questions) and full (36 questions) versions in persons with low back pain

Author	Population				Instrument Administration						
	N	Age (y), Mean \pm SD (Range)	Sex (% Female)	Health Condition, Duration, Mean \pm SD, Proportion With LBP (%)	Setting	Country, Who Region, Low-/Middle-/High-Income Country*	Version of Instrument, Language	Participation Rate, Follow-up Rate (If Applicable)	Mode of Administration [†]	Scoring Method [‡]	Comparator Instrument (If Applicable)
Chronic LBP Bärlund et al ¹²	1379	47.6 \pm 14.9	64.0	Chronic LBP 92.0% had a main diagnosis of M - Diseases of the MSK system and connective tissue Most frequent diagnosis M54-Dorsalgia in 50.0% Chronic duration not specified	Outpatient physical and rehabilitation medicine clinic at a university hospital	Finland European region High-income	WHODAS 2.0 short version Language not specified	Participation rate 44.0%	Self-administered	Summary scores Simple scoring 0-48	ODI
Cwirlej-Sozanska et al ¹³	92	66.0 \pm 11.6	62.0	Chronic LBP (\geq 12wk)	Hospital rehabilitation ward	Poland European region High-income	WHODAS 2.0 full version Polish	Participation rate study I: 87.6% Follow-up for study II: conducted 2 d after 94.6% Follow-up for study III: conducted 1 mo after completing rehabilitation and leaving the hospital 74.7%	Direct interview implementing a pen and paper interview method (self-reported questionnaire)	Summary and domain scores Complex scoring 0-100 scale	SF-36, ODI, HADS, VAS for pain
Gonçalves Silva et al ³⁷	504	70.9 \pm 7.5	67.1	Various chronic conditions; LBP (n=206) 54.8% LBP duration <6 mo (n=16) 7.7%, \geq 6 mo (n= 190) 92.3%	Primary health care practices	Portugal European region High-income	WHODAS 2.0 short version Portuguese	Participation rate not reported	Interview-based	Summary scores Simple scoring 0-60	SPPB (for performance-based disability)
Igwesi-Chidobe et al ¹⁵	12 (cross-cultural adaptation, pilot/pretesting sample) 50 (test-retest reliability sample) 200 (Construct validity sample)	45.0 \pm 10.36; 45.2 \pm 11.55; 48.6 \pm 12.0	41.7; 64.0; 44.0	Chronic LBP Chronic duration not specified	Urban and rural areas of Nigeria	Nigeria African region Lower-middle income	WHODAS 2.0 full version Igbo	Participation rate not reported	Interview-based	Summary and domain scores Complex scoring 0-100 scale	Igbo-RMDQ, BPS, BS-11
Other Fujiwara et al ¹⁴	369	70.5 \pm 10.0	50.7	Any pain. The most common diseases resulting in pain were spondylosis and spinal stenosis Duration not specified	Pain clinic at Nara Medical University	Japan Western Pacific Region High-income	WHODAS 2.0 short version Japanese	Participation rate 99.7%	Self-reported questionnaire	Summary score Complex scoring 0-100 scale	NRS, EQ-5D5L, TMIG Index (IADL, IA, SR)
Garin et al ³⁸	1190	52.7 \pm 15.6	56.2	Multiple chronic conditions; LBP (n=118) 9.9% Duration not specified	Seven different European centers	Czech Republic, Germany, Italy, Slovenia, Spain European Region High-income	WHODAS 2.0 full version Language not specified	Participation rate not reported for entire sample LBP specific, stable at 6 weeks n=17 LBP specific, improved at 3 mo n=11	Self-administered or interviewer-administered	Summary and domain scores Complex scoring 0-100 scale	NRS for pain; Terciles: Mild (0-3), Moderate (4-6), Severe (7-10)
Katajapuu et al ¹⁷	1988	47.6 \pm 15.0	65.0	Chronic MSK pain 88.0% had a main diagnosis of M - Diseases of the MSK system and connective tissue. Most frequent diagnoses M54 Dorsalgia in 39.0% Chronic duration not specified	Outpatient physical and rehabilitation medicine clinic of a university hospital	Finland European region High-income	WHODAS 2.0 short version Language not specified	Participation rate 63.0%	Self-administered questionnaire	Summary score Simple scoring 0-48 scale (also presented as a percentage of 100%)	NA

(continued on next page)

Table 2 (Continued)

Author	N	Population			Instrument Administration						
		Age (y), Mean \pm SD (Range)	Sex (% Female)	Health Condition, Duration, Mean \pm SD, Proportion With LBP (%)	Setting	Country, Who Region, Low-/Middle-/High-Income Country*	Version of Instrument, Language	Participation Rate, Follow-up Rate (If Applicable)	Mode of Administration†	Scoring Method‡	Comparator Instrument (If Applicable)
Katajapuu et al ¹⁸	1988	47.6 \pm 15.0	65.0	Chronic MSK pain 88.0% had a main diagnosis of M - Diseases of the MSK system and connective tissue. Most frequent diagnoses M54 Dorsalgia in 39.0% Chronic duration not specified	Outpatient physical and rehabilitation medicine clinic of a university hospital	Finland European Region High-income	WHODAS 2.0 short version Language not specified	Participation rate 63.0%	Self-administered questionnaire	Summary scores Simple scoring 0-48 scale (also presented as a percentage of 100%)	NA
Katajapuu et al ¹⁶	1988	47.6 \pm 6.3	65.0	Chronic MSK pain 88.0% had a main diagnosis of M - Diseases of the MSK system and connective tissue. Most frequent diagnoses M54 Dorsalgia in 39.0% Chronic duration not specified	Outpatient physical and rehabilitation medicine clinic of a university hospital	Finland European region High-income	WHODAS 2.0 short version Language not specified	Participation rate 63.0%	Self-administered questionnaire	Summary score Simple scoring 0-48 scale	NA
Pösl et al ³⁹	296	54.0 \pm 11.3	50.0	MSK conditions (LBP, rheumatoid arthritis, osteoarthritis) % with LBP not specified Duration not specified	Inpatient or outpatient rehabilitation (multicenter study)	Germany European region High-income	WHODAS 2.0 full version German	Participation rate not reported	Patient self-administration form of WHODAS-36	Summary and domain scores Complex scoring 0-100 scale	SF-36
Røe et al ⁴⁰	118	47.0 \pm 12.0	52.0	Acute LBP (<4wk), subacute LBP (4-12wk) or chronic LBP (>12wk) 13.0% acute 11.0% subacute 76.0% chronic	4 study centers	Norway European region High-income	WHODAS 2.0 full version Norwegian	Participation rate not reported	Self-reported questionnaire	Domain and item scores Simple scoring 0-5 scale per item	Comprehensive ICF Core Set for LBP, ODI
Saltychev et al ^{41,42}	501	47.0 \pm 13.7 (range, 16-84)	65.0	MSK pain conditions 36.0% with back pain (ICD-10 code M54) Other relevant ICD-10 codes: 11.0% with Intervertebral Disk Disorder (M51), 2.0% with Dorsopathy (M53), 2.0% with Deforming Dorsopathy (M43), 2.0% with Nerve root or plexus disorder (G54), 2.0% with Spondylopathy (M48) Duration not specified	Outpatient physical and rehabilitation medicine clinic of a university hospital	Finland European region High-income	WHODAS 2.0 short version Finnish	Participation rate 67.0%	Self-administered questionnaire	Summary scores Simple scoring 0-48 scale (also presented as a percentage of 100%)	NA
Silva et al ⁴³	204	65.9 \pm 9.1	71.6	Individuals experiencing pain linked to MSK pathologies (eg, osteoarthritis 67.2%, spondylarthrosis 46.6%) Duration of pain not specified	Rehabilitation clinics	Portugal European region High-income	WHODAS 2.0 full version Portuguese	Participation rate not reported	Interview-based	Summary and domain scores Complex scoring 0-100	Visual analog scale
Tarvonen-Schröder et al ¹⁹	81	47.2 \pm 9.5	63.0	Chronic spinal pain 61.7% with lumbosacral pain Chronic duration not specified	Rehabilitation ward of a tertiary clinic in a university hospital	Finland European region High-income	WHODAS 2.0 short version Language not specified	Participation rate not reported	Self-administered questionnaire (patient or proxy)	Summary and domain scores Simple scoring 0-48 scale	WHO minimal generic data set

Abbreviations: BPS, Back Performance Scale; BS-11, 11-point Box Scale; EQ-5D5L, EuroQol-5 Dimension 5-Level; HADS, Hospital Anxiety and Depression Scale; IADL, instrumental independence; IA, intellectual mobility; ICD-10, *International Classification of Diseases, Tenth Revision*; Igbo-RMDQ, Igbo-Roland-Morris Disability Questionnaire; MSK, musculoskeletal; NA, not applicable; NRS, numeric rating scale; ODI, Oswestry Disability Index; SPPB, Short Physical Performance Battery tests; SR, social role; TMIG Index, Tokyo Metropolitan Institute of Gerontology Index; VAS, visual analog scale.

* As per the World Bank.

† Mode of administration includes self-report, interview-based, parent/proxy report, and so on.

‡ Scoring method includes scale, summary or domain score, simple or complex scoring.

19.1±6.8), and persons with no pain (mean, 14.5±3.6),³⁷ suggesting adequate construct validity.

One study reported correlations (significant correlation coefficient >0.3 considered valid) but did not explicitly state hypothesis related to direction of correlations; thus, construct validity is indeterminate. In persons with musculoskeletal conditions (including LBP), correlation between WHODAS 2.0 short version summary scores and (1) numeric rating scale for pain was $r=0.37$ (95% CI, 0.27-0.45); (2) EuroQoL 5-Dimension 5-Level was $r=-0.66$ (95% CI, -0.72 to -0.60); and (3) Tokyo Metropolitan Institute of Gerontology Index summary score was $r=-0.67$ (95% CI, -0.72 to -0.61).¹⁴ Another study did not provide any a priori hypotheses, so construct validity is indeterminate. In persons with chronic musculoskeletal conditions (including LBP), correlation between WHODAS 2.0 short version summary scores and Oswestry Disability Index was $r=0.71$ ($P<.05$).¹²

Cross-cultural validity/measurement invariance

One study investigated the sex-related differential item function of WHODAS 2.0 short version in persons with chronic musculoskeletal conditions (including LBP).¹⁸ No significant differences in WHODAS 2.0 short version summary scores (simple scoring) were reported between men (mean, 27.2±19.9) and women (mean, 27.4±19.3). Statistically significant differential item function between sexes was observed in 7 of the 12 items. Detected differential item functions were uniform, suggesting that the direction of sex-related differences between responses persisted across the spectrum of disability severity. Men had to report slightly worse disability on WHODAS 2.0 short version items “household,” “emotional affection,” and “work” to achieve the same score as women. In contrast, women had to report slightly worse disability on WHODAS 2.0 short version items “concentration,” “washing,” “dressing,” and “dealing with strangers” to achieve the same score as men.

Internal consistency

Evidence from 6 studies suggests that WHODAS 2.0 full^{13,15,39,43} and short^{14,16} versions have adequate internal consistency in persons with LBP (see tables 3 and 4).

Test-retest and interrater reliability

Evidence from 3 studies suggests that WHODAS 2.0 full version has sufficient test-retest and interrater reliability in persons with LBP (see tables 3 and 4).^{13,15,43}

Minimal detectable change

Four studies estimated the MDC for WHODAS 2.0 full^{13,15,43} and short¹⁶ versions (see tables 3 and 4). For WHODAS 2.0 full version summary scores, MDC was 12.33 of 100 using simple scoring,⁴³ and ranged from 10.45-13.99 of 100 for complex scoring.^{13,15} The MDC was 8.6 of 48 for WHODAS 2.0 short version summary scores using simple scoring.¹⁶

Responsiveness

Evidence from 2 studies suggests that WHODAS 2.0 full version may have adequate responsiveness in persons with LBP^{13,39} (see table 3). In persons with chronic LBP who completed a

rehabilitation program, mean change in WHODAS 2.0 full version summary scores (complex scoring) was -11.97 ± 8.86 of 100 (effect size= -0.86).¹³ In persons with musculoskeletal conditions of mixed duration (including LBP) who completed a rehabilitation program, mean change in WHODAS 2.0 full version summary scores (complex scoring) was -4.7 of 100 (SD and effect size not reported).³⁹

Floor and ceiling effects

Evidence from 3 studies suggests no floor or ceiling effects with WHODAS 2.0 full version summary scores^{13,15} but potential floor effects with WHODAS 2.0 short version summary scores¹⁷ in persons with LBP (appendices 7A and 7B). In persons with chronic LBP, there were no floor or ceiling effects in WHODAS 2.0 full version summary scores on a 0-100 scale (complex scoring), which were defined as $\geq 15\%$ of participants providing the lowest or highest possible score on WHODAS 2.0 full version, respectively.^{13,15} In persons with chronic musculoskeletal conditions (including LBP), a significant floor effect (set at $>15\%$) was observed for WHODAS 2.0 short version summary scores (simple scoring, 0-48 scale), but no ceiling effects were observed.¹⁷

Minimal important change

Two studies estimated the MIC of WHODAS 2.0 full version in persons with chronic LBP¹³ and of the short version in persons with chronic musculoskeletal conditions (including LBP)¹⁶ (see appendices 7A and 7B). In persons with chronic LBP, MIC of WHODAS 2.0 full version (complex scoring) was estimated as 4.87 of 100 or 9.74 of 100 (corresponding to 1- or 2-point change on 0-10 cm visual analog scale for pain intensity, respectively), calculated using a linear regression model.¹³ In persons with chronic LBP, MIC of WHODAS 2.0 short version (simple scoring) was estimated as 3.09 of 48 (using $0.33\times SD$), 3.10 of 48 (using SEM), and 4.68 of 48 (using $0.5\times SD$), calculated using distribution-based methods.¹⁶

Discussion

Evidence suggests that WHODAS 2.0 full version has adequate content and structural validity in persons with LBP, but construct validity is indeterminate. Evidence suggests that WHODAS 2.0 full version has adequate internal consistency, test-retest reliability, and interrater reliability in persons with LBP. For WHODAS 2.0 short version, evidence suggests that it has adequate structural validity and internal consistency in persons with LBP, but construct validity is indeterminate. The MDC ranged from 10.45-13.99 of 100 for WHODAS 2.0 full version and 8.6 of 48 for the short version. We found no floor or ceiling effects with the full version but potential floor effects with the short version scores. MIC for WHODAS 2.0 full version was estimated at 4.87 of 100 or 9.74 of 100 (corresponding to 1- and 2-point change on 0-10 cm visual analog scale for pain intensity, respectively) and ranged from 3.09-4.68 of 48 for the short version.

Our systematic review provides more comprehensive and up-to-date evidence on the measurement properties and MIC of WHODAS 2.0 in measuring functioning in persons with LBP than previous reviews.^{10,11} The systematic review by Federici et al reported that WHODAS 2.0 correlated with disease-specific measures in persons with back pain based on 2 studies and did not find

Table 3 Evidence table on select measurement properties of WHODAS 2.0 short (12 questions) and full (36 questions) versions in persons with low back pain

Author; WHODAS 2.0 Version	Internal Consistency	Cross-Cultural Validity/ Measurement Invariance	Reliability	Measurement Error	Construct Validity	Responsiveness
Chronic LBP						
Bärlund et al ¹² ; WHODAS 2.0 short version	NA	NA	NA	NA	Correlation between WHODAS total score and ODI total score $r=0.71$ $P<.0001$	NA
Cwirlej-Sozanska et al ¹³ ; WHODAS 2.0 full version	Cronbach α 0.92 (for total score)	NA	ICC _{2,1} (95% CI): 0.928 (0.898- 0.949) for the total score Time between the 2 measurements amounted to 2 d on average and were made by different interviewers	For WHODAS total score: SEM: 3.77 MDC ₉₅ : 10.45	Correlation between WHODAS total score and SF-36 domains $r=-0.349$ to -0.873 , $P<.001$ Correlation between WHODAS total score and HADS domains anxiety $r = 0.455$, $P<.001$ and HADS domain depression $r=0.489$, $P<.001$ Correlation between WHODAS total score and ODI $r=0.867$, $P<.001$ WHODAS total score for Pain VAS 0-4, mean \pm SD, 27.06 \pm 6.85; $P<.001$ WHODAS total score for Pain VAS 5-10, mean \pm SD, 45.29 \pm 12.68; $P<.001$	For WHODAS total score: SRM: -1.35 ES: -0.86 SE: 0.24 Change between first and third study of WHODAS total score: mean \pm SD, -11.97 ± 8.86 ; median, -11.96
Gonçalves Silva et al ³⁷ ; WHODAS 2.0 short version	NA	NA	NA	NA	Comparing participants with LBP with those with pain reported elsewhere or no pain: For LBP group: mean global pain intensity 5.6 \pm 2.6 (using vertical numeric graphic rating scale) For LBP group: WHODAS total score mean \pm SD, 23.4 \pm 8.7; $P<.01$ (when compared with the no pain group) Individuals who reported pain elsewhere: WHODAS total score mean \pm SD, 19.1 \pm 6.8; $P<.05$ (when compared with the no pain group)	NA
Igwesi-Chidobe et al ¹⁵ ; WHODAS 2.0 full version	Cronbach α 0.97 (for total score)	NA	ICC (95% CI): 0.93 (0.88- 0.96) for the total score For test-retest reliability, measures completed at baseline and repeated 7-10 d post baseline by the same community health worker	For WHODAS total score: SEM: 5.05 MDC: 13.99	Correlation between Igbo-WHODAS total score and Igbo-RMDQ $r=0.54$, $P<.01$ Correlation between Igbo-WHODAS total score and BPS $r=0.34$, $P<.01$ Correlation between Igbo-WHODAS total score and BS-11 $r=0.56$, $P<.01$	NA
Other						
Fujiwara et al ¹⁴ ; WHODAS 2.0 short version	Cronbach α 0.91 (entire scale)	NA	NA	NA	Correlation between WHODAS and NRS $r=0.37$ (95% CI, 0.27-0.45) $P<.001$ Correlation between WHODAS and EQ-5D5L $r=-0.66$ (95% CI, -0.72 to -0.60) $P<.001$ Correlation between WHODAS and TMIG Index, total $r=-0.67$ (95% CI, -0.72 to -0.61) $P<.001$ Correlation between WHODAS and TMIG Index, IADL $r=-0.64$ (95% CI, -0.69 to -0.57) $P<.001$ Correlation between WHODAS and TMIG Index, IA $r=-0.40$ (95% CI, -0.48 to -0.31) $P<.001$ Correlation between WHODAS and TMIG Index, SR $r=-0.53$ (95% CI, -0.59 to -0.44) $P<.001$	NA
Garin et al ³⁸ ; WHODAS 2.0 full version	NA	NA	NA	NA	No statistically significant difference between WHODAS total score LBP mild group (NRS 0-3) vs LBP severe group (NRS 7-10), ES 0.45	NA
Katajapuu et al ¹⁸ ; WHODAS 2.0 short version	NA	Subgroups: men, women No difference in summary score between men (27.2 \pm 19.9) and women (27.4 \pm 19.3); $P=.843$ Significant ($P<.05$) uniform DIF: Men who endorsed item had slightly higher WHODAS total score than women for "household responsibilities," "being emotionally affected," "work" Women who endorsed item had slightly higher WHODAS than men for "concentrating for 10 minutes," "washing," "dressing," "dealing with strangers" WHODAS-12 items dichotomized as "none" (rated by respondents as 0) vs "any limitation" (rated by respondents as 1, 2, 3, or 4)	NA	NA	NA	NA

(continued on next page)

Table 3 (Continued)

Author; WHODAS 2.0 Version	Internal Consistency	Cross-Cultural Validity/ Measurement Invariance	Reliability	Measurement Error	Construct Validity	Responsiveness
Katajapuu et al ¹⁶ ; WHODAS 2.0 short version	Cronbach α 0.89 (entire scale)	NA	NA	For WHODAS total score: SEM: 3.10 MDC: 8.6 Minimal detectable percentage change 66%	NA	NA
Pösl et al ³⁹ ; WHODAS 2.0 full version	Cronbach α 0.87 (WHODAS understanding and communicating) Cronbach α 0.85 (WHODAS getting around) Cronbach α 0.82 (WHODAS self-care) Cronbach α 0.70 (WHODAS getting along with others) Cronbach α 0.94 (WHODAS activities) Cronbach α 0.82 (WHODAS participation)	NA	NA	NA	Correlation between WHODAS total score and SF-36 mental health -0.43 , SF-36 vitality -0.50 , SF-36 pain -0.53 , SF-36 physical functioning -0.62 , SF-36 role physical -0.49 , SF-35 role emotional -0.40 , SF-36 social functioning -0.60 ; $P < .01$ for all of the above	Mean \pm SD for WHODAS total score before rehabilitation intervention: 22.0 ± 14.3 Mean \pm SD for WHODAS total score after rehabilitation intervention: 17.3 ± 16.1
Røe et al ⁴¹ ; WHODAS 2.0 full version	NA	NA	NA	NA	NA	NA
Saltychev et al ^{41,42} ; WHODAS 2.0 short version	NA	NA	NA	NA	NA	NA
Silva et al ⁴³ ; WHODAS 2.0 full version	Cronbach α 0.84 (for total score) Only 65 responses included in this analysis of the 204 due to nonapplicable responses in domain 5 "activities of daily living" by the rest of the participants	NA	ICC (95% CI): 0.95 (0.94-0.96) for total score For inter-evaluator reliability, the WHODAS-36 was administered in 2 different interviews by 2 different interviewers Minimum and maximum time interval of 1-3 d	For WHODAS total score: *SEM: 4.45 *MDC: 12.33	WHODAS score between participants with widespread pain: mean \pm SD, 35.7 ± 13.3 WHODAS score between participants with pain in only 1 spot: mean \pm SD, 24.1 ± 28.4 ; $P = .009$ WHODAS score between participants with pain in 2 or 3 sites: mean \pm SD, 29.0 ± 14.1 ; $P = .003$ Correlation between WHODAS and intensity of pain (visual numeric scale) $P < .01$	NA
Tarvonen-Schröder et al ¹⁹ ; WHODAS 2.0 short version	NA	NA	NA	NA	Correlation between WHODAS patient ($n=81$) and proxy ($n=31$) sum scores in patients with chronic spinal pain $r=0.78$, $P < .0001$ Correlation between WHODAS sum scores (patient and proxy) and WHO minimal generic data set $r=0.48$, $P < .0001$ WHODAS-12 scores rated by patients with chronic spinal pain, mean \pm SD, 22.0 ± 0.99 compared with spinal cord injury, mean \pm SD, 18.4 ± 1.05 ; $P < .05$	NA

Abbreviations: BPS, Back Performance Scale; BS-11, 11-point Box Scale; DIF, differential item functioning; EQ-5D5L, EuroQoL-5 Dimension 5-Level; ES, effect size; HADS, Hospital Anxiety and Depression Scale; IADL, instrumental independence; IA, Intellectual mobility; Igbo-Roland-Morris Disability Questionnaire; NA, not applicable; NRS, numeric rating scale; ODI, Oswestry Disability Index; SR, social role; SRM, standardized response mean; TMIG Index, Tokyo Metropolitan Institute of Gerontology Index; VAS, visual analog scale.

* Calculated by authors of the systematic review.

any studies reporting on MIC for WHODAS 2.0.¹⁰ The systematic review by Saltychev et al identified 3 studies assessing the structural validity and floor and ceiling effects of WHODAS 2.0 short version in persons with musculoskeletal pain. Similar to our review findings, Saltychev et al reported a 2-factor structure, with a floor effect but no ceiling effect with the short version.¹¹ The systematic reviews were limited by a narrow search strategy that did not include subject headings and absence of critical appraisal to assess the methodological quality of studies.^{10,11} These systematic reviews also needed updating because the literature search conducted by Federici et al¹⁰ spanned 1995-2015 and that by Saltychev et al ended January 2019.¹¹ Our review adds new evidence

on the construct validity of WHODAS 2.0 based on 10 studies and found that other aspects of validity, reliability, and responsiveness are adequate for WHODAS 2.0 in persons with LBP. Moreover, we synthesized more recently published studies that estimated the MIC of WHODAS 2.0 in persons with LBP.^{13,16}

Our review findings have implications for clinical practice, population surveillance, and future research. WHODAS 2.0 questionnaire comprehensively measures functioning across major domains because it is directly linked to the ICF framework. Given adequate measurement properties, health care providers and decision makers can consider using WHODAS 2.0 questionnaire to measure functioning in persons with LBP. In clinical settings,

Table 4 Evidence on internal consistency, reliability, and minimal detectable change for WHODAS 2.0 short (12 questions) and full (36 questions) versions in persons with low back pain

Measurement Property	Population	Questionnaire	Key Results
Internal consistency	Chronic LBP	WHODAS 2.0 full version summary scores	Ranged from Cronbach $\alpha=0.92$ ¹³ - 0.97 ¹⁵
	Musculoskeletal conditions including LBP	WHODAS 2.0 full version summary scores	Cronbach $\alpha=0.84$ ⁴⁴
		WHODAS 2.0 full version domain scores	Cronbach $\alpha=0.70$ (domain for getting along with others) to $\alpha=0.94$ (domain for activities) for domain scores ³⁹
	Musculoskeletal conditions of mixed duration including LBP	WHODAS 2.0 short version summary scores	Cronbach $\alpha=0.89$ ¹⁶ - 0.91 ¹⁴
Test-retest reliability	Chronic LBP	WHODAS 2.0 full version summary scores	Between 2 measurements taken 2 d apart: ICC=0.93 (95% CI, 0.90-0.95). ¹³
			Between 2 measurements taken 7-10 d apart: ICC=0.93 (95% CI, 0.88-0.96) ¹⁵
Interrater and test-retest reliability	Musculoskeletal conditions of mixed duration including LBP	WHODAS 2.0 full version summary scores	Administered again via interview by a second interviewer within 3 d: ICC 0.95 (95% CI, 0.94-0.96) ⁴³
Minimal detectable change	Chronic LBP	WHODAS 2.0 full version summary scores (complex scoring)	Ranged from 10.45/100 ¹³ to 13.99/100 ¹⁵
	Musculoskeletal conditions including LBP	WHODAS 2.0 full version summary scores (simple scoring)	12.33/100 ⁴⁴
	Chronic musculoskeletal conditions including LBP	WHODAS 2.0 short version summary scores (simple scoring)	8.6/48 ¹⁶

health care providers can use WHODAS 2.0 to measure functioning to guide management, to inform prognosis, and as an appropriate outcome measure for assessing rehabilitation for LBP. At the population level, WHODAS 2.0 can be used in population-based surveys to measure functioning related to LBP across cultures and settings, particularly the version with 12 questions because it is shorter and easy to administer. Future research is needed to estimate the MIC of WHODAS 2.0 short and full versions in persons with LBP using anchor-based methods. We found 2 studies that estimated the MIC of WHODAS 2.0, 1 study that used linear regression and 1 that used distribution-based methods to calculate MIC.^{13,16} Guidance from COSMIN recommends that the MIC be determined using an anchor-based longitudinal approach with multiple anchors to reflect a value that patients consider important.²² Studies are also needed to assess the construct validity of WHODAS 2.0 short and full versions, using a priori hypotheses that state the direction and strength of correlations for hypothesis testing.²² In addition, we did not find any relevant studies targeting subacute LBP (6-week to 3-month duration), so future research assessing the measurement properties of WHODAS 2.0 in this population is warranted.

The MDC is estimated to be larger than the MIC for WHODAS 2.0 in persons with LBP. This can pose as a challenge to the utility of the instrument, and varying approaches to address this issue are discussed in the literature.⁴⁴⁻⁴⁸ We suggest that changes in functioning measured using WHODAS 2.0 may be considered as achieving both thresholds indicated by MIC and MDC to be meaningful because they represent different concepts in measurement.^{45,49} This includes an estimated MDC of 10.45-13.99 of 100 and MIC of 9.74 of 100 for WHODAS 2.0 full version (complex scoring) as well as MDC of 8.6 of 48 and MIC of 4.68 of 48 for the short version (simple scoring). In addition, health care providers, decision makers, and other knowledge users using WHODAS 2.0 may consider

supplementing with an additional question to assess for self-rated improvement to facilitate applying an anchor-based approach.⁴⁴

Study strengths and limitations

Our systematic review has strengths and offers unique contributions to this area. Our systematic review was planned a priori with a registered protocol on Open Science Framework Registries and reported based on the Preferred Reporting Items for Systematic Reviews and Meta-analyses statement.²⁴ Our search strategy was comprehensive and reviewed by a second librarian using the Peer Review of Electronic Search Strategies Checklist.³³ In addition, we used the recommended COSMIN and COSMIN-OMERACT checklists to assess risk of bias in studies on measurement properties and MIC of patient-reported questionnaires.^{23,34} We also calculated the MDC of WHODAS 2.0 based on available data in studies.

Our systematic review has limitations. First, we included studies published in English, French, Farsi, Chinese, German, Greek, Swedish, or Italian languages to increase feasibility. However, a previous study found no evidence of systematic bias when using language restrictions in systematic reviews in conventional medicine.⁵⁰ We did not identify any possibly relevant studies published in other languages during screening. Second, we recognize that subjectivity and judgment are involved in critical appraisal and took several methodological steps to minimize its potential effect. Specifically, we (1) trained reviewers to screen and critically appraise studies using standardized methodology, (2) used a standardized critical appraisal tool to assess risk of bias (COSMIN and COSMIN-OMERACT checklists),^{23,34} (3) provided detailed notes that accompanied the COSMIN and COSMIN-OMERACT tools and edited to match the purpose of this review, and (4)

undertook a quality control assessment with an independent reviewer to review all critical appraisal checklists, risk of bias ratings, and extracted data.

As a limitation of the current state of literature, we found few published studies assessing certain measurement properties of each version of the WHODAS 2.0 (eg, cross-cultural validity/measurement invariance, MIC) for LBP; further research is warranted. In addition, most studies were conducted in high-income countries, including the European region (79% of included studies). Therefore, some results may not be generalizable to other settings, including low- or middle-income countries.

Conclusions

Evidence suggests that WHODAS 2.0 full version has adequate content validity, structural validity, internal consistency, and reliability in persons with LBP. WHODAS 2.0 short version has adequate structural validity and internal consistency in persons with LBP. The construct validity of the short and full versions is indeterminate. The MIC has been estimated for WHODAS 2.0 in persons with LBP. Since MDC is estimated to be larger than MIC, health care providers and decision makers may consider using both MIC and MDC thresholds to measure change in functioning in persons with LBP. Our review findings have implications for health care providers, decision makers, and other knowledge users who intend to measure functioning in persons with LBP. Health care providers can consider using WHODAS 2.0 to measure functioning to guide management, to inform prognosis, and as a rehabilitation outcome measure for LBP. At the population level, WHODAS 2.0 can be considered for use in population-based surveys to measure functioning in persons with LBP across cultures and settings.

Author contributions

Conceptualization: J.W., A.D., S.H.J., W.D.G., A.C., P.C. Methodology: J.W., A.D., S.H.J., W.D.G., D.S., M.B., N.L., S.A., S.M., P.S., M.N., P.C. Literature search: A.T.V. Data collection: J.W., A.D., D.S., M.B., N.L., S.A., H.V., P.C. Writing – original draft: J.W., with input from A.D., P.C. Writing – review and editing: J.W., A.D., S.H.J., W.D.G., D.S., M.B., N.L., S.A., H.V., S.M., P.S., M.N., A.T.V., A.C., P.C. All authors read and approved the final manuscript.

Keywords

Low back pain; Rehabilitation; World Health Organization; Surveys and questionnaires; Back pain; International classification of functioning, disability and health

Corresponding author

Pierre Côté, DC, PhD, Institute for Disability and Rehabilitation Research, Ontario Tech University, 2000 Simcoe St North, Oshawa, Ontario, Canada L1G 0C5. *E-mail address:* pierre.cote@ontariotechu.ca.

Acknowledgments

We thank Dr Dorcas Beaton, PhD, for helpful suggestions on methodology and Mr Kent Murnaghan, MA, MIST, for peer review of the search strategy.

References

- Hoy D, March L, Brooks P, et al. The global burden of low back pain: estimates from the Global Burden of Disease 2010 study. *Ann Rheum Dis* 2014;73:968–74.
- Wu A, March L, Zheng X, et al. Global low back pain prevalence and years lived with disability from 1990 to 2017: estimates from the Global Burden of Disease Study 2017. *Ann Transl Med* 2020;8:299.
- Hartvigsen J, Hancock MJ, Kongsted A, et al. What low back pain is and why we need to pay attention. *Lancet* 2018;391:2356–67.
- Schofield DJ, Shrestha RN, Passey ME, et al. Chronic disease and labour force participation among older Australians. *Med J Aust* 2008;189:447–50.
- Carey TS, Evans A, Hadler N, et al. Care-seeking among individuals with chronic low back pain. *Spine (Phila Pa 1976)* 1995;20:312–7.
- Carey TS, Evans AT, Hadler NM, et al. Acute severe low back pain. A population-based study of prevalence and care-seeking. *Spine (Phila Pa 1976)* 1996;21:339–44.
- GBD 2017 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet* 2018;392:1789–858.
- Cieza A, Causey K, Kamenov K, et al. Global estimates of the need for rehabilitation based on the Global Burden of Disease study 2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet* 2020;396:2006–17.
- World Health Organization. Measuring health and disability: manual for World Health Organization (WHO) Disability Assessment Schedule 2.0 (WHODAS 2.0). 2012. Available at: <https://www.who.int/standards/classifications/international-classification-of-functioning-disability-and-health/who-disability-assessment-schedule>. Accessed April 23, 2021.
- Federici S, Bracalenti M, Meloni F, et al. World Health Organization Disability Assessment Schedule 2.0: an international systematic review. *Disabil Rehabil* 2017;39:2347–80.
- Saltychev M, Katajapuu N, Bärlund E, et al. Measurement properties of 12-item self-administered World Health Organization Disability Assessment Schedule 2.0 (WHODAS 2.0) among general population and people with non-acute physical causes of disability - systematic review. *Disabil Rehabil* 2021;43:789–94.
- Bärlund ET, Katajapuu NK, Paltamaa JP, et al. Correlation between Oswestry Disability Index and 12-item self-administered version of World Health Organization Disability Assessment Schedule (WHODAS 2.0) in patients with chronic low back pain. *Int J Rehabil Res* 2021;44:170–2.
- Ćwirlej-Sozańska A, Bejer A, Wiśniewska-Szurlej A, et al. Measurement properties of the Polish Version of the 36-Item WHODAS 2.0 in patients with low back pain. *Int J Environ Res Public Health* 2020;17:7284.
- Fujiwara A, Ida M, Kimoto K, et al. Feasibility, reliability, and validity of the 12-item World Health Organization Disability Assessment Schedule 2.0 in patients attending the pain clinic. *J Anesth* 2021;35:81–5.
- Igwesi-Chidobe CN, Kitchen S, Sorinola IO, et al. World Health Organisation Disability Assessment Schedule (WHODAS 2.0): development and validation of the Nigerian Igbo version in patients with chronic low back pain. *BMC Musculoskelet Disord* 2020;21:755.
- Katajapuu N, Heinonen A, Saltychev M. Minimal clinically important difference and minimal detectable change of the World Health Organization Disability Assessment Schedule 2.0 (WHODAS 2.0) amongst

- patients with chronic musculoskeletal pain. *Clin Rehabil* 2020;34:1506–11.
17. Katajapuu N, Laimi K, Heinonen A, et al. Floor and ceiling effects of the World Health Organization Disability Assessment Schedule 2.0 among patients with chronic musculoskeletal pain. *Int J Rehabil Res* 2019;42:190–2.
 18. Katajapuu N, Laimi K, Heinonen A, et al. Gender-related differences in measurement properties of WHO Disability Assessment Schedule 2.0. *Int J Rehabil Res* 2019;42:316–21.
 19. Tarvonen-Schröder S, Kaljonen A, Laimi K. Comparing functioning in spinal cord injury and in chronic spinal pain with two ICF-based instruments: WHODAS 2.0 and the WHO minimal generic data set covering functioning and health. *Clin Rehabil* 2019;33:1241–51.
 20. Nicol R, Yu H, Selb M, et al. How does the measurement of disability in low back pain map unto the International Classification of Functioning, Disability and Health?: a scoping review of the manual medicine literature. *Am J Phys Med Rehabil* 2021;100:367–95.
 21. Wang D, Taylor-Vaisey A, Negrini S, et al. Criteria to evaluate the quality of outcome reporting in randomized controlled trials of rehabilitation interventions. *Am J Phys Med Rehabil* 2021;100:17–28.
 22. Mokkink LB, Prinsen C, Patrick DL, et al. COSMIN methodology for systematic reviews of patient-reported outcome measures (PROMs). User manual 2018:78. Available at: https://www.cosmin.nl/wp-content/uploads/COSMIN-syst-review-for-PROMs-manual_version-1_feb-2018-1.pdf. Accessed August 1, 2021.
 23. Outcome Measures in Rheumatology. COSMIN-OMERACT good methods checklist. Available at: <https://omeract.org/instrument-selection/downloadable-forms/>. Accessed June 30, 2021.
 24. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021;372:n71.
 25. Duthey B. World Health Organization Background Paper 6.24. Low back pain. Available at: http://www.who.int/medicines/areas/priority-medicines/BP6_24LBP.pdf. Accessed August 1, 2018.
 26. Salehi R, Negahban H, Khiavi FF, et al. Validity and reliability of the World Health Organization Disability Assessment Schedule 2.0 36-Item Persian Version for persons with multiple sclerosis. *Korean J Fam Med* 2020;41:195–201.
 27. Cheung MK, Hung AT, Poon PK, et al. Validation of the World Health Organization Assessment Schedule II Chinese Traditional Version (WHODAS II CT) in persons with disabilities and chronic illnesses for Chinese population. *Disabil Rehabil* 2015;37:1902–7.
 28. Higgins AM, Neto AS, Bailey M, et al. The measurement properties and minimal clinically important difference for disability assessment using WHODAS 2.0 in critically ill patients. *Crit Care Resusc* 2021;23:103–12.
 29. Küçükdeveci AA, Kutlay Ş, Yıldızlar D, et al. The reliability and validity of the World Health Organization Disability Assessment Schedule (WHODAS-II) in stroke. *Disabil Rehabil* 2013;35:214–20.
 30. Park SH, Demetriou EA, Pepper KL, et al. Validation of the 36-item and 12-item self-report World Health Organization Disability Assessment Schedule II (WHODAS-II) in individuals with autism spectrum disorder. *Autism Res* 2019;12:1101–11.
 31. Silveira C, Souza RT, Costa ML, et al. Validation of the WHO Disability Assessment Schedule (WHODAS 2.0) 12-item tool against the 36-item version for measuring functioning and disability associated with pregnancy and history of severe maternal morbidity. *Int J Gynaecol Obstet* 2018;141(Suppl 1):39–47.
 32. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol* 2010;63:737–45.
 33. McGowan J, Sampson M, Salzwedel DM, et al. PRESS Peer Review of Electronic Search Strategies: 2015 guideline statement. *J Clin Epidemiol* 2016;75:40–6.
 34. Mokkink LB, de Vet HCW, Prinsen CAC, et al. COSMIN Risk of Bias checklist for systematic reviews of patient-reported outcome measures. *Qual Life Res* 2018;27:1171–9.
 35. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J Strength Cond Res* 2005;19:231–40.
 36. Campbell M, McKenzie JE, Sowden A, et al. Synthesis without meta-analysis (SWiM) in systematic reviews: reporting guideline. *BMJ* 2020;368:16890.
 37. Silva AG, Queirós A, Rocha NP. Generic self-reported and performance based instruments: how to capture pain associated disability. *Rev Port Saúde Pública* 2016;34:125–33.
 38. Garin O, Ayuso-Mateos JL, Almansa J, et al. Validation of the "World Health Organization Disability Assessment Schedule, WHODAS-2" in patients with chronic diseases. *Health Qual Life Outcomes* 2010;8:1–15.
 39. Pösl M, Cieza A, Stucki G. Measurement properties of the WHODAS-II in rehabilitation patients. *Qual Life Res* 2007;16:1521–31.
 40. Røe C, Sveen U, Bautz-Holter E. Retaining the patient perspective in the International Classification of Functioning, Disability and Health Core Set for low back pain. *Patient Prefer Adherence* 2008;2:337–47.
 41. Saltychev M, Bärlund E, Mattie R, et al. A study of the measurement properties of 12-item World Health Organization Disability Assessment Schedule 2.0 in a large population of people with chronic musculoskeletal pain. *Clin Rehabil* 2017;31:262–72.
 42. Saltychev M, Mattie R, McCormick Z, et al. Confirmatory factor analysis of 12-Item World Health Organization Disability Assessment Schedule in patients with musculoskeletal pain conditions. *Clin Rehabil* 2017;31:702–9.
 43. Silva C, Coleta I, Silva AG, et al. Adaptation and validation of WHODAS 2.0 in patients with musculoskeletal pain. *Rev Saude Publica* 2013;47:752–8.
 44. de Vet HC, Terwee CB. The minimal detectable change should not replace the minimal important difference. *J Clin Epidemiol* 2010;63:804–5. [author reply 806].
 45. de Vet HC, Terwee CB, Ostelo RW, et al. Minimal changes in health status questionnaires: distinction between minimally detectable change and minimally important change. *Health Qual Life Outcomes* 2006;4:54.
 46. Terwee CB, Roorda LD, Dekker J, et al. Mind the MIC: large variation among populations and methods. *J Clin Epidemiol* 2010;63:524–34.
 47. Terwee CB, Roorda LD, Knol DL, et al. Linking measurement error to minimal important change of patient-reported outcomes. *J Clin Epidemiol* 2009;62:1062–7.
 48. Turner D, Schünemann HJ, Griffith LE, et al. The minimal detectable change cannot reliably replace the minimal important difference. *J Clin Epidemiol* 2010;63:28–36.
 49. de Vet HC, Beckerman H, Terwee CB, et al. Definition of clinical differences. *J Rheumatol* 2006;33:434. [author reply 435].
 50. Morrison A, Poliseña J, Husereau D, et al. The effect of English-language restriction on systematic review-based meta-analyses: a systematic review of empirical studies. *Int J Technol Assess Health Care* 2012;28:138–44.