Validity and reliability of instruments for global measurement of disability: Systematic literature review

Validez y fiabilidad de instrumentos para la medición global de la discapacidad: revisión sistemática de la literatura

Camila Rodríguez-Guevara[1] ; Andry Yasmid Mera-Mamián[1] ; Diana Isabel Muñoz-Rodríguez[1] ;
Gino Montenegro-Martínez[1] ; Pablo Esteban Roa-Urrutia[2] ; Erika Alejandra Giraldo-Gallo[1]

* camilarodriguevara1@gmail.com

Abstract

Introduction: Various theoretical and conceptual frameworks have historically shaped the measurement of disability, many focusing on activity limitations due to impairments in bodily functions or structures. However, these perspectives do not comprehensively address other components, such as environmental factors and participation in life situations under various health conditions, which would provide a more holistic measurement of global disability. **Objective:** This study aimed to provide a comprehensive overview of the available instruments for assessing disability using the ICF model. Additionally, the study sought to examine the validity and reliability of assessment procedures applied to these instruments. **Methodology:** A systematic literature review was conducted following PRISMA guidelines. Searches were performed in Ovid, Embase, LILACS, Scopus, Rehabilitation Reference Center, and Google Scholar from 2012 to 2022. Independent reviewers performed screening, selection, and data extraction. Risk of bias assessment according to COSMIN and GRADE level of evidence was applied. **Results:** A total of 1,998 articles were identified, 188 were reviewed in full text, and 3 were included in the review. The identified scales for assessing global disability were WHODAS 2.0 and IMPACT-S. The quality of measurement properties for the first scale was indeterminate for structural validity and internal consistency and sufficient for hypothesis testing; the level of evidence was moderate. The IMPACT-S was indeterminate for structural validity and sufficient for internal consistency, reliability, criterion validity, and hypothesis testing. **Conclusions:** The most widely used instrument for measuring global disability is WHODAS 2.0, which has proven preventable across different contexts and populations.

Keywords: Disability; Disability evaluation; Adults; Self-testing; Psychometrics; Validation study; Reproducibility of results.

[1] Universidad CES, Medellín, Colombia
[2] Secretaría Distrital de Santiago de Cali. Cali, Colombia.

SALUD UIS

Resumen

Introducción: históricamente, la medición de discapacidad obedece a diferentes perspectivas teóricas y conceptuales, gran parte de las cuales identifican las limitaciones en actividades concretas generadas por una deficiencia en una función o estructura corporal. Sin embargo, no abordan de forma holística otros componentes como los factores ambientales y la participación en situaciones vitales en diferentes condiciones de salud en general, lo cual brindaría una medición completa de la discapacidad global como propone la Clasificación Internacional del Funcionamiento y la Discapacidad (CIF). **Objetivo:** proporcionar una revisión completa de los instrumentos disponibles para evaluar la discapacidad global según la CIF. Además, se buscó examinar la validez y fiabilidad de los procedimientos de evaluación aplicados a estos instrumentos. **Metodología:** Revisión sistemática de la literatura de acuerdo con criterios de PRISMA. Se realizó búsqueda en Ovid, Embase, LILACS, Scopus, Rehabilitation Reference Center y Google Scholar en el período de 2012 a 2022. El tamizaje, la selección y extracción de los datos se realizó por evaluadores independiente. Se aplicó la evaluación del riesgo de sesgos acorde a COSMIN y nivel de evidencia de GRADE. **Resultados:** se identificaron 1998 artículos, se revisaron 188 a texto completo y se incluyeron 3 en la revisión. Las escalas identificadas para evaluar discapacidad global fueron WHODAS 2.0 e IMPACT-S. El nivel de evidencia de acuerdo con GRADE fue moderada para el cuestionario WHODAS 2.0. Respecto al IMPACT-S fue indeterminada la validez estructural y suficiente para la consistencia interna, fiabilidad, validez de criterio y test de hipótesis. **Conclusiones:** el instrumento más empleado para medir discapacidad global es el WHODAS 2.0 el cual ha demostrado ser válido y confiable en diferentes contextos y poblaciones.

Palabras clave: Discapacidad; Evaluación de la discapacidad; Adultos; Autoevaluación; Psicometría; Estudio de validación; Reproducibilidad de los resultados.

Introduction

Disability is a multifaceted notion that can be treated from several angles and for various objectives[1]. Some authors argue that there is no single idea of disability, but somewhat multiple sorts of disabilities[2], whereas others advocate for the presentation of a coherent concept[1]. These conceptual distinctions are related, among other things, to a shift in their social and health approaches. Initially, the medical model focused on understanding and analyzing the deficits or deficiencies of body structures that resulted in disability[3]. In contrast, the social model influences the contemporary vision, which emphasizes the importance of the individual's interaction with his environment[4].

The World Health Organization (WHO) incorporates the biopsychosocial model into its International Classification of Functioning, Disability and Health (ICF) to define disability and functioning as a dynamic interaction among health conditions, contextual factors, and personal factors[5]. Furthermore, disability is commonly regarded as a comprehensive concept encompassing impairments in anatomical structures and physiological functions, restrictions in performing activities, and limitations with engaging in various life[6]. The assessment of disability presents an enormous challenge due to the multifaceted nature of the phenomenon. Nonetheless, it is crucial to scrutinize socioeconomic measurements and apportion resources to gauge its prevalence[7]. Historically, self-report measures and performance-based assessments have been used, with a greater emphasis on the former[8]. This is because self-report measures allow for a better understanding of the individual's perception instead of only depending on professional observation[9].

A wide range of self-report instruments are available, some focused on risk factors, others on prognosis, and the majority aimed at detecting deficiencies in bodily structures and functions[10,11]. The WHODAS 2.0 questionnaire is a prevalent tool in various fields of study. The World Health Organization Disability Assessment Schedule, released by the WHO in 2010, uses a set of 36 questions to comprehensively evaluate an individual's functioning and disability on a global scale. The measure's validity and reliability have been demonstrated across diverse cultural contexts and among populations with varying health conditions[12-15].

Given the significance of staying abreast of the latest developments in disability research and the ongoing challenge of achieving consistency in its measurement, this study aims to comprehensively overview the available instruments for assessing disability using the ICF model. Additionally, the study sought to examine the validity and reliability of assessment procedures applied to these instruments[16].

Method

Registries and search techniques

PROSPERO registered the protocol (CRD42022348222). This study adheres to PRISMA[17] and COSMIN[18] guidelines. Medline (Ovid), Embase, LILACS, Scopus, Rehabilitation Reference Center, Virtual Health Library (VHL), and Google Scholar were searched for information. Based on the research query (**Appendix 1**), the search terms were determined using the Descriptors in Health Sciences (DeCs) in Spanish and the Medical Subject Headings (MeSH) in English. In addition, accessible terms were included (**Appendix 2**). As Boolean operators, "OR" and "AND" were used. The search was limited to English, Spanish, and German publications only. The study was restricted to studies published between 2012 and 2022.

Criteria for eligibility and study selection

Studies were done on people over the age of 18 with a variety of health conditions and levels of severity. This review centered on questionnaires, surveys, profiles, and other tools clearly labeled as self-reporting tools in the studies. These tools are called PROMs or "Patient-reported outcome measures." Based on the COSMIN guideline[18], the important measurement values for the review are shown in **Table 1**. Instruments that measured outcomes other than disability, such as functional capacity, quality of life, health status, and return to work, were excluded. Also, those that didn't measure disability solely focused on a specific body function or structure or were made to measure disability in groups with certain diseases.

Table 1. Evaluation of measurement properties on included items.

Content validity
Development of the PROM (Patient Reported Outcome Measurement)
Content validity
Internal structure
Structural validity
Internal Consistency
Cross-Cultural Validity/Measurement of Invariance
Remaining Measurement Properties
Reliability
Measurement error
Criterion validity
Hypothesis test for construct validity
Responsiveness

Source: Authors

Two phases were involved in selecting the studies. In the first step, three researchers (DM, PR, and CR) independently reviewed the titles and abstracts of the papers based on the inclusion and exclusion criteria. In the event of a disagreement, the matter was resolved by an agreement between researchers. In the second phase, the complete texts of the selected articles were independently reviewed by three reviewers (DM, PR, and CR). The articles included in the review were selected from the second review (AM, CR). The open-access platform Rayyan Systematic Review[19] was used for the articles section.

Information extraction

Two reviewers (CR and AM) conducted the information extraction independently. Microsoft 365 Excel was used to create a matrix containing information determined beforehand regarding the general characteristics of the studies and the characteristics of the instruments. Concerning the characteristics of the studies, data were extracted on the PROM, the study population (sample size, age, and gender), and the health condition. The following details about the instruments were extracted: validation language, domains, mode and time of administration, recall period, evaluated properties, and quantitative and qualitative results. Researchers (AM and CR) convened to review each article and verify the extracted data according to the predefined matrix. To ensure the quality of extracted information in instances of disagreement, the article was directly reviewed to validate the data and reach a consensus on the precise information relevant to the study.

Methodological quality evaluation

The evaluation of the quality of the methodology was conducted by two evaluators independently (AM and with COSMIN proposal[18]. First, each study's risk of bias was evaluated using the COSMIM checklist (18). In addition, four hypotheses were developed to assess the construct validity of the instruments. Each hypothesis and its supporting theory are presented in the following table:

In general, the population with mental health issues (such as anxiety and depression) has a higher overall disability score than individuals without this background. The research reports that moderate disability is evident in people with depression and physical pain, with scores of 37.7±19.2[20] and 24.77±23.00[21], respectively. At the same time, the disability is only modest for those who have an anxiety illness, with a score of 17.28±8.98[22].

Based on the literature review, the following hypothesis was formulated regarding the application of the WHODAS 2.0 questionnaire: in people with significant depression and bodily pain, the score obtained would indicate moderate disability, while in people with anxiety disorder, a mild disability would be expected. A good to excellent correlation between the WHODAS 2.0 questionnaire and IMPACT-S is also anticipated.

There is a moderate to strong negative correlation ≥ 0.50 between the disability score and quality of life; the more significant the disability, the lesser the quality of life[18].

There is a moderate to strong positive correlation ≥ 0.50 between various instruments used to assess disability[18]. The quality criteria were then applied to every measurement property. Using the *Grading of Recommendations, Assessment, Development, and Evaluation-GRADE model*, the level of evidence was estimated for scales that were analyzed in more than

one study. In **Appendix 4** describes the assessment of risk of bias in a study on structural validity, internal consistency, criterion validity, reliability, and hypotheses testing for construct validity. Additionally, **Appendix 5** contains the criteria for generic hypotheses to evaluate construct validity and responsiveness.

Results

Literature search

A total of 1,998 articles were initially collected, but after removing duplicates, 1,766 were subjected to title and abstract review. Of these, 1,588 articles were eliminated, leaving 188 for full-text review. Following this, 185 articles were excluded for not satisfying the eligibility criteria, resulting in identifying three studies that met the requirements for the systematic review[12,23,24]. In **Appendix 3**, the search strategy for each database is described. The flowchart presented in **Figure 1** depicts searching and selecting articles for the study.
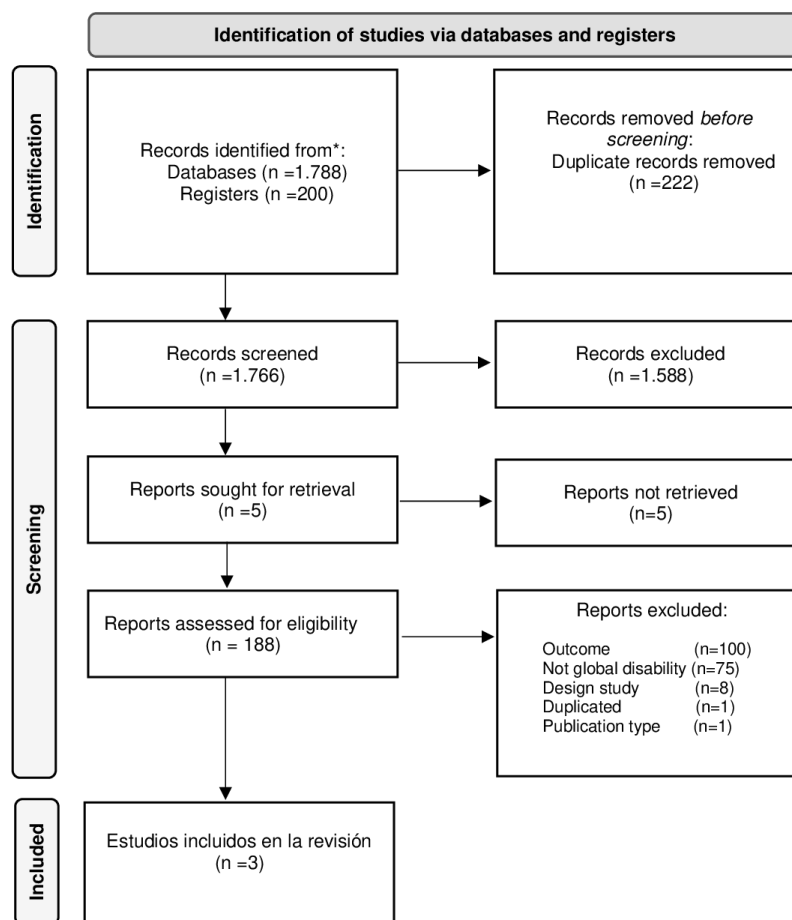

Figure 1. PRISMA 2020 flow diagram for new systematic reviews which included searches of databases and registers only.

Characteristics of the included studies

Three validation studies[12,23,24] with a total of 308 participants were included; 136 (44.2%) had multiple sclerosis, 79 (25.6%) had head trauma, and 93 (30.2%) had cerebrovascular disease. Two studies reporting the WHODAS 2.0[12,23] and one survey of IMPACT-S[24] were analyzed. The characteristics of the included inquiries are presented in **Table 2**.

WHODAS 2.0 Questionnaire

General characteristics of the studies

Magistrale et al.[12] included 136 multiple sclerosis patients. 80.9% had relapsing-remitting forms of multiple sclerosis, while 19.1% had progressive forms. The participants' mean age was 42.94 ± 11.18 years (range: 19–72). The preponderance was female (71.3%)[12]. Snell et al. conducted a preliminary validation of the short version of 12 questions in 79 adult patients with a diagnosis of moderate CIE-10 traumatic brain injury due to being struck by an object (33%), motor vehicle accidents (29%), or falls (28%). The average age was 41,5 years, and 56% were women[23].

Psychometric properties assessed in research studies

Magistrale et al.[12] identified a mean score of 18.2±16.1. Rasch's analysis indicated that despite redundant items, the 36-item and 32-item scales fit the model well (PSI = 0.83).

It was only able to solve four out of seven subscales during the analysis (a high incidence of extreme scores). Cronbach's Alpha was more significant than 0.79 for most subscales and 0.93 for the total scale, indicating acceptable to excellent reliability based on the classical model. After eliminating two items on intimate relationships and sexual activity, the subscale for "get along with others" obtained a score of 0.68.

A significant correlation was found between MSQoL-54 (Multiple Sclerosis Quality of Life-54) composite scores and WHODAS 2.0 subscales regarding convergent validity. For instance, the MSQoL-54 physical health composite showed a strong correlation (>0.70) with the score of the WHODAS 2.0 "move" and "participation" subscales. This final analysis was conducted with only 53 participants from another sub study on domestic accidents[12].

In the study by Snell et al.[23], WHODAS 2.0 was administered via telephone follow-up. Internal consistency was measured with a Cronbach's alpha of 0.92. Similarly, they conducted an exploratory factor analysis employing the Kaiser-Meyer-Olkin (KMO) measure, varimax rotation, and Bartlett's sphericity test with a p-value of 0.001. This version contains three factors; the first explains 53.1% of the variance and relates to questions regarding novel learning tasks, community activities, relationships with other people, and work activities. By adding the second factor associated with questions about domestic chores, maintaining prolonged positions, and walking, 64.0% of the variance is explained, and 72.6% of the variance is presented by the third factor, which inquiries about self-care activities.

Additional analyses revealed that it also has adequate discriminant validity, as it is capable of distinguishing statistically significant differences between patients with comorbidities such as major depression, anxiety disorder, and bodily pain and those without comorbidities (p < 0.001)[23]. The characteristics of the reported measurement properties are listed in **Table 3**.

Table 2. Characteristics of included studies

Instruments (Author, year)	Country	Population and Sample	Population Characteristics	Health conditions
WHODAS II (Magistrale et al. 2015)	Italy	136 adults	42.94 ± 11,18 years (range 19–72) 39 men (28.7%)	Long-lasting multiple sclerosis. Mean MG 10,5 ± 10,4 years
WHODAS II (Snell et al. 2017)	Vancouver, Canada	79 adults	Adults 18 to 65 years 41.5 years ± 12.0 Range (19-64 years) Men 35 (44.3%)	Cranioencephalic trauma (TEC)
IMPACT-S (Schenk et al. 2015)	Germany	93 adults	Adults 27 to 90 years 62,1±12,1 years Men 47 (50,1%)	Cerebrovascular event

Source: Authors

Quality of the measurement properties and level of evidence

Magistrale et al. and Snell et al. evaluated three out of ten measurement properties. As the study by Snell included fewer than 100 participants in the Factorial Analysis, structural validity received an overall indeterminate rating for risk of bias assessment. Snell et al. compiled a global report, not one for each domain, leaving the internal consistency indeterminate. The construct validity hypothesis test received an adequate rating because the studies by Magistrale et al. and Snell et al. supported the reviewers' hypotheses[12,23]. According to the COSMIN criteria, if 75% of the hypotheses are supported, they are deemed adequate[18].

Lastly, the GRADE level of evidence was moderate because the risk of bias in the study by Snell et al. was high due to a questionable structural validity rating and inadequate internal consistency (previously stated arguments). However, the quality of the study by Magistrale et al. was adequate, and the results are consistent with Cronbach's Alpha values in the range of 0.92 to 0.93; it is not possible to assess consistency for the remaining properties because they are not discovered in both studies. Given that the sample size is > 200 participants (n=236), the results are accurate. In conclusion, the evidence is pertinent to the review's research query. **Table 3** provides a summary of the results of this procedure.

IMPACT-S questionnaire

General characteristics of the studies

Schenk et al. validated this questionnaire in German among 93 patients diagnosed with Cerebrovascular Events. Patients in the initial stages of rehabilitation were not recruited. The average age was 62.1 years, 49.5% were female, and the most prevalent diagnosis was stroke. Additionally, 53.1% received a pension, and the average level of education was above average[24].

Psychometric properties assessed in research studies

Other authors developed the process of cultural and trans-linguistic adaptation in 2011. The global internal consistency was 0.97, ranging from 0.70 to 0.90 for the individual domains, which is outstanding. The observation window for test-retest reliability was 7 to 14 days, and the Intraclass Correlation Coefficient (ICC) ranged from 0.77 to 1.00, which is excellent to flawless. With a Spearman correlation coefficient of 0.79, the construct validity revealed a strong relationship between the activity and participation subscales.

The Principal Component Analysis (PCA) of the 9 theoretically defined domains yielded two components, the first explaining 5.9% and 65.7% of the explained variation and the second explaining 11.9%. The first aspect was related to task domains and general requirements, communication, learning, and knowledge application. The WHODAS 2.0 questionnaire was used for concurrent validity, with a Spearman correlation coefficient of 0.85[24]. **Table 4** summarizes the reported measurement properties' features.

Quality of the measurement properties and level of evidence

The authors assessed 4 out of 10 measurement qualities. According to the risk of bias evaluation, the structural validity needed to be improved due to a sample size of 5 participants per item. Similarly, they did not include the adjusted comparative index or the Turkey-Lewis coefficient. When a Cronbach's Alpha of 0.70 was obtained, the internal consistency, in evaluating the danger of bias, and its quality were classified as very good and sufficient, respectively. Despite a questionable risk of bias since it was unclear whether the same application conditions were maintained, the test-retest reliability was acceptable to reach an ICC of 0.70. Concurrent validity with WHODAS 2.0 (the gold standard) was assessed as "very good and sufficient" with a Spearman's correlation of 0.70. The hypothesis test for the validity of the convergent construct was adequate because it achieved a correlation of more than 0.50; similarly, it complied with the reviewers' hypotheses by being between 0.65 and 0.80. The outcomes of this method are summarized in **Table 4**.

Table 3. Synthesis of the quality of the measurement properties and level of evidence of the WHODAS 2.0 questionnaire.

Author	Content validity	Structural Validity	Internal Consistency	Cross-Cultural Validity/ Invariance	Reliability	Error measurement	Criterion validity	Hypothesis Test for construct validity	Responsiveness	Level of evidence
(Magistrale, et. al, 2015)	____	A	VG	____	____	____	____	VG	____	
(Snell, et.al. 2017)	____	D	I	____	____	____	____	VG	____	Moderate
Overall quality criterion	____	?	?	____	____	____	____	+	____	

NOTE. VG = Very Good; A = Adequate; I= Inadequate; D = Doubtful; + = Enough;? = Indeterminate; D.A.= Does not apply.
Source: Authors

Table 4. Synthesis of the quality of the measurement properties and level of evidence of the IMPACT-S questionnaire.

Author	Content validity	Structural Validity	Internal Consistency	Cross-Cultural Validity/ Invariance	Reliability	Error measurement	Criterion validity	Hypothesis Test for construct validity	Responsiveness	Level of evidence
(Schenk, 2015)	____	I	VG	____	D	____	MB	MB	____	D.A.
Overall quality criterion	____	?	+	____	+	____	+	+	____	

NOTE. VG = Very Good; A = Adequate; I= Inadequate; D = Doubtful; + = Enough? = Indeterminate; D.A.= Does not apply.
Source: Authors

Discussion

A comprehensive disability requires evaluating the individual within their environment to determine the impact of their condition on overall performance[25]; thus, measures that focus disability on a specific structure, function, or activity were not included. In this review, it is believed that they have significant conceptual constraints. The widely used WHODAS 2.0 with cross-cultural applicability and the lesser-known IMPACT-S created by Post et al. in 2008[26] were included. According to the author's study, the instruments were verified in a population with chronic health issues, and between three and four measurement properties were investigated. According to the methodological quality assessment, the GRADE level of evidence for WHODAS 2.0 was moderate. The IMPACT-S scale is a trustworthy and valid instrument.

WHODAS 2.0 information on validity and reliability was published in 2010. According to the research, the structural validity was assessed using exploratory and confirmatory factor analysis, and a two-level hierarchical structure was established, with a general factor of disability feeding the six domains[25]. The Snell study in this review revealed three parameters, albeit the sample size utilized (100) for the investigating feature was deemed questionable[23]. In the case of Magistrale, he

discovered that the scales of 36 and 32 items obtained an excellent adjustment to the Rasch model (PSI = 0.83) despite redundant elements and the inability to resolve four out of seven in the analysis by subscales. (high rate of extreme scores), which is why it advises that future analyses incorporate larger samples and revisit the instrument's initial structure and length[12].

They showed internal consistency with item-total values ranging from "acceptable" to "very good"[25] in the WHODAS 2.0 handbook, although Magistrale[12] and Snell[23] found good-excellent findings with a higher overall score of 0.92. The Snell investigation showed no evidence of internal consistency by factor. Even though WHODAS measures something else, the previously established concurrent validity between WHODAS 2.0 and instruments such as the London Disability Scale (LHS), SF-36 Health Survey, and Functional Independence Measure (FIM), among others, correlates with these dimensions[25]. Only Magistrale evaluated this attribute in this review, using the MSQoL-54, a self-report instrument that measures patients' quality of life with multiple sclerosis. A high association was discovered with the "move" and "participation" subscales, indicating that it has external validity[12].

Snell also assessed the instrument's discriminant validity, or its ability to detect differences in disability

in subpopulations with specific characteristics and discovered that the highest scores were for subgroups of patients with comorbidities, particularly those with post-contusion syndrome U = 221.00; Z = 5.12, $p < 0.01$, Cohen's d = 1.37). It was, however, discovered in individuals with major depressive illness U = 160.50; Z = 5.23; p < 0.01; Cohen's d = 1.69) and anxiety (U = 302.00 [Z = 4.23]; p < 0.01; Cohen's d = 1.13)[23]. Previously, it was discovered that the World Mental Health Survey, which employed the initial edition of WHODAS, had a discriminating capability and a strong impact on mental health status, psychological and physical problems, job incapacity, and quality of life[27-28].

Other analyses that Magistrale or Snell did not include are reported in the official WHODAS 2.0 publication. For example, in face validity, they discovered that 64% of professionals believed the WHODAS 2.0 evaluates impairment, according to the ICF. However, this concept is always changing as the policies surrounding it change. They reported an ICC of 0.69-0.89 at the question level, 0.93-0.96 at the domain level, and 0.98 at a general level, which, while adequate, should be considered when examining the instrument in populations with unique criteria[25].

Post et al. developed the IMPACT-S questionnaire in the Netherlands in 2008. It is a generic instrument that is also based on the ICF and considers nine domains of life that are developed in 33 items. It comprises two sub-scores (activity and participation) and a total score ranging from 0 to 100, with a lower score indicating less disability[26]. The principal component analysis revealed two factors, the first accounting for the variance and the second 12.2%, a finding similar to Schenk's[24]. Poust, conversely, indicates that, according to the analyses, the characteristics did not show a separation between activity and involvement[29].

Regarding internal consistency, Post et al.[29] found a correlation of 0.96 for the total score[26] and Schenk et al. of 0.97, interpreted as excellent[24]. The original study's test-retest reliability was 0.58, according to the Kappa Index[26]. However, Schenk et al. used the ICC and found a higher result between 0.77 and 1.00[24]. Concurrent validity with WHODAS 2.0 showed a high correlation between each IMPACT-S domain and the total score, being 0.88[26], very similar to that reported by Schenk et al. with a Spearman correlation of 0.85. that is, from good to excellent[24].

One of the review's drawbacks is the modest number of included articles that met the qualifying criterion. The improper identification of the goods under the Patient-Reported Outcome Measures (PROM) label, which is the technical and controlled designation for this equipment, could lead to this predicament. According to the search results, most are found under the scale's name in the development or validation process. Another restriction is the inability to do a meta-analysis due to the variability of the populations analyzed and the validation procedures used. Furthermore, due to a lack of information on the measurement parameters, the evaluation of publication bias still needs to be performed, even though it is recommended to be done with a minimum of 10 papers[30].

Conclusions

Although widely disseminated in clinical, academic, and research disciplines, global disability by ICF concepts is not yet fully represented across scales, questionnaires, or other assessment tools. On the other hand, the WHODAS 2.0 questionnaire is the most developed instrument in its validation and reliability process in populations with various health conditions, whereas the IM. In contrast, it shows a different level of development despite being based on the ICF. Finally, this research is novel as it evaluates certainty using the GRADE methodology in studies that report the psychometric properties of global disability questionnaires.

Authors contribution

All authors contributed substantially to the design, conception, collection, analysis, or interpretation of data, preparing or critically revising the manuscript and approving the work's final version..

Acknowledge

Ethical considerations

The research follows ethical considerations and scientific integrity by registering the protocol on the PROSPERO platform and adequately citing the studies in the review.

Conflict of Interest

The authors declare that there are no conflicts of interest.

Funding

This systematic review did not receive financial funding.

AI technological support

The authors report that they did not use Artificial Intelligence, language models, machine learning or similar technologies to create or assist with the elaboration or editing of any of the contents of this document.

References

1. Timpe K. Denying a unified concept of disability. J Med Philos. 2022; 47(5). doi: https://doi.org/10.1093/jmp/jhac021

2. Jeff M. Radical cognitive limitation. In disability and disadvantage. In: Brownlee K, Cureton A, editors. Disability and disadvantage. Londres: Oxford University Press; 2009. p. 240–59.

3. McDermott S, Turk MA. The myth and reality of disability prevalence: Measuring disability for research and service. Disabil Health J. 2011; 4(1): 1-5. doi: https://doi.org/10.1016/j.dhjo.2010.06.002

4. Braddock D, Parish S. Disability at the dawn of the 21st Century and The State of the States. Quinta Edición. Braddock D, editor. Washington: American Association on Mental Retardation; 2002. 1–86 p.

5. Organización Mundial de la Salud, Organización Panamericana de la Salud. Clasificación internacional del funcionamiento de la discapacidad y de la salud. Primera Edición. Organización Mundial de la Salud, Organización Panamericana de Salud, editors. Washington D. C: OPS; 2001. 1–258 p.

6. 1O'Young B, Gosney J, Ahn C. The concept and epidemiology of disability. Phys Med Rehabil Clin N Am. 2019; 30(4): 697-707. doi: https://doi.org/10.1016/j.pmr.2019.07.012

7. Bourke JA, Nichols-Dunsmuir A, Begg A, Dong H, Schluter PJ. Measuring disability: An agreement study between two disability measures. Disabil Health J. 2021; 14(2): 100995. doi: https://doi.org/10.1016/j.dhjo.2020.100995

8. Manini T. Development of physical disability in older adults. Curr Aging Sci. 2011; 4(3): 184-191. doi: https://doi.org/10.2174/1874609811104030184

9. Sen A. Health: perception versus observation. BMJ. 2002; 324(7342): 860-861. doi: https://doi.org/10.1136/bmj.324.7342.860

10. Stuck AE, Walthert JM, Nikolaus T, Büla CJ, Hohmann C, Beck JC. Risk factors for functional status decline in community-living elderly people: A systematic literature review. Soc Sic Me. 1999; 48(4): 445-469. doi: https://doi.org/10.1016/S0277-9536(98)00370-0

11. Tas Ü, Verhagen AP, Bierma-Zeinstra SMA, Odding E, Koes BW. Prognostic factors of disability in older people: A systematic review. Br J Gen Pract. 2007; 57(537): 319-323.

12. Magistrale G, Pisani V, Argento O, Incerti CC, Bozzali M, Cadavid D, et al. Validation of the world health organization disability assessment schedule II (WHODAS-II) in patients with multiple sclerosis. Mult Scler. 2015; 21(4): 448–456. doi: https://doi.org/10.1177/1352458514543732

13. Meesters JJL, Verhoef J, Liem ISL, Putter H, Vlieland TPMV. Validity and responsiveness of the World Health Organization Disability Assessment Schedule II to assess disability in rheumatoid arthritis patients. Rheumatology (Oxford). 2010; 49(2): 326-333. doi: https://doi.org/10.1093/rheumatology/kep369

14. Wolf T, Schulz H, Losem C, Reichert D, Hurtz HJ, Sandner R, et al. Prophylaxis of chemotherapy-induced neutropenia and febrile neutropenia with lipegfilgrastim in patients with non-Hodgkin lymphoma (NADIR study). Eur J Haematol. 2019; 102(2): 174–181. doi: https://doi.org/10.1111/ejh.13189

15. Zacarias LC, Câmara KJ da C, Alves BM, Morano MTAP, Viana CMS, Mont'Alverne DGB, et al. Validation of the World Health Organization Disability Assessment Schedule (WHODAS 2.0) for individuals with COPD. Disabil Rehabil. 2022; 44(19): 5663-5668. doi: https://doi.org/10.1080/09638288.2021.1948117

16. O'Young B, Gosney J, Ahn C. The concept and epidemiology of disability. Phys Med Rehabil Clin N Am. 2019; 30(4): 697–707. doi: https://doi.org/10.1016/j.pmr.2019.07.012

17. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. Declaración PRISMA 2020: una guía actualizada para la publicación de revisiones sistemáticas. Rev Esp Cardiol. 2021; 74(9): 790-799. doi: https://doi.org/10.1016/j.recesp.2021.06.016. Erratum in: Rev Esp Cardiol (Engl Ed). 2022; 75(2): 192.

18. Mokkink LB, Prinsen CA, Patrick DL, Alonso J, Bouter LM, de Vet HC, et al. COSMIN methodology for systematic reviews of patient-reported outcome measures (PROMs). COSMIN manual for systematic reviews of PROMs COSMIN. 2018. Available from: https://cosmin.nl/wp-content/uploads/COSMIN-syst-review-for-PROMs-manual_version-1_feb-2018.pdf.

19. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev. 2016; 5(1): 210. doi: https://doi.org/10.1186/s13643-016-0384-4

20. Chiang YC, Liou TH, Lee HC, Escorpizo R. Using whodas 2.0 to assess functional impairment in people with depression: Should employment receive more attention? Int J Environ Res Public Health. 2021; 18(9): 4552. doi: https://doi.org/10.3390/ijerph18094552

21. Ćwirlej-Sozańska A, Sozański B, Kotarski H, Wilmowska-Pietruszyńska A, Wiśniowska-Szurlej A. Psychometric properties and validation of the polish version of the 12-item WHODAS 2.0. BMC Public Health. 2020; 20(1): 1230. doi: https://doi.org/10.1186/s12889-020-09305-0

22. Hoehne A, Giguère CE, Herba CM, Labelle R. Assessing functioning across common mental disorders in psychiatric emergency patients: Results from the WHODAS-2. Canad J Psychiatry. 2021; 66(12): 1085-1093. doi: https://doi.org/10.1177/0706743720981200

23. Snell DL, Iverson GL, Panenka WJ, Silverberg ND. Preliminary validation of the World Health Organization Disability Assessment Schedule 2.0 for mild traumatic brain injury. J Neurotrauma. 2017; 34(23): 3256–3261. doi: https://doi.org/10.1089/neu.2017.5234

24. Schenk Zu, Schweinsberg E, Lange J, Schucany M, Wendel C. Teilhabe nach Schlaganfall - Validierung der deutschen Übersetzung des IMPACT-S [Participation Following Stroke - Validation of the German Version of IMPACT-S]. Rehabilitation (Stuttg). 2015; 54(3): 160-165. German. doi: https://doi.org/10.1055/s-0035-1545358

25. Organización Mundial de la Salud. Medición de la Salud y la Discapacidad - Manual para el Cuestionario de Evaluación de la Discapacidad de la OMS. Organización Mundial de la Salud (OMS). 2015. Disponible en: https://apps.who.int/iris/bitstream/handle/10665/170500/9874573309_spa.pdf?sequence=1

26. Post MWM, de Witte LP, Reichrath E, Verdonschot MM, Wijlhuizen GJ, Perenboom RJM. Development and validation of impact-s, an ICF-based questionnaire to measure activities and participation. J Rehabil Med. 2008; 40(8): 620-627. doi: https://doi.org/10.2340/16501977-0223

27. Alonso J, Angermeyer MC, Bernert S, Bruffaerts R, Brugha TS, Bryson H, et al. Disability and quality of life impact of mental disorders in Europe: Results from the European study of the epidemiology of mental disorders (ESEMeD) project. Acta Psychiatr Scan Suppl. 2004; 109(420): 38-46. doi: https://doi.org/10.1111/j.1600-0047.2004.00329.x

28. Buist-Bouwman MA, Ormel J, De Graaf R, Vilagut G, Alonzo G, Van Sonderen E, et al. Psychometric properties of the World Health Organization Disability Assessment Schedule used in the European Study of the Epidemiology of Mental Disorders. Int J Methods Psychiatr Res. 2008; 17(4): 185-197. doi: https://doi.org/10.1002/mpr.261

29. Post MWM, van der Zee CH, Hennink J, Schafrat CG, Visser-Meily JMA, van Berlekom SB. Validity of the utrecht scale for evaluation of rehabilitation-participation. Rehabil. 2012; 34(6): 478–485. doi: https://doi.org/10.3109/09638288.2011.608148

30. Centro Cochrane Iberoamericano traductores. Manual Cochrane de revisiones sistemáticas de intervenciones versión 5.1.0. versión 5.1.0. Higgins J, Green S, editors. 2012. 1–639 p.

Appendix. Supplementary Materials

Validity and reliability of instruments for global measurement of disability: Systematic Literature Review

Appendix 1. Review question of Patient-Reported Outcome Measures (PROMs)

Structure review question	
C **(Construct of interest)**	Self-Assessment; Surveys and Questionnaires; Disability Evaluation
P **(Population)**	Population: young adult; adult; middle Aged; aged; aged, 80 and over.
T **(Type of measurement instrument)**	Patient-Reported Outcome Measures (PROM)
M **Measurement Properties**	Psychometrics; Psychometric properties; Reproducibility of Results

Question:
What are the instruments that, according to their psychometric properties, are recommended for measuring disability and functioning in people over 18 years of age?

Appendix 2. Keywords

Structure Question	DECs terms	MESH terms	Meaning	Synonyms
C	Auto-evaluación	Self-Assessment	Appraisal of one's own personal qualities or traits	Assessment, Self Self Assessment Self-Criticism
	Cuestionario	Surveys and Questionnaires	Collections of data obtained from voluntary subjects. The information usually takes the form of answers to questions, or suggestions.	Questionnaire Questionnaires Respondent Respondents Survey Surveys
	Evaluación de discapacidad	Disability Evaluation	Determination of the degree of a physical, mental, or emotional handicap. The diagnosis is applied to legal qualification for benefits and income under disability insurance and to eligibility for Social Security and workmen's compensation benefits.	No tiene sinónimos
P	Joven adulto	Young adult	For a person between 19 and 24 years of age.	No tiene
	Adulto	Adult	Adults are of 19 through 44 years of age.	Adults
	Adulto	Middle Aged	An adult aged 45 - 64 years.	Middle Age
	Adulto mayor	Aged	A person 65 through 79 years of age	Elderly
	Adulto mayor	AGED, 80 AND OVER	For a person older than 79 years	Elderly

Structure Question	DECs terms	MESH terms	Meaning	Synonyms
T	Instrumentos de resultado informadas por el paciente (PROM)	Patient Reported Outcome Measures (PROM)	Assessment of the quality and effectiveness of health care as measured and directly reported by the patient.	Patient Reported Outcome Patient Reported Outcome Measure Patient Reported Outcomes Patient-Reported Outcome Patient-Reported Outcomes
M	Psicometría	"psychometrics"	Assessment of psychological variables by the application of mathematical procedures.	No tiene
	Propiedades psicométricas	Psychometric properties (Término libre)	Assessment of psychological variables by the application of mathematical procedures.	No tiene
	Reproducibilidad de resultados	Reproducibility of Results	The statistical reproducibility of measurements (often in a clinical context), including the testing of instrumentation or techniques to obtain reproducible results. The concept includes reproducibility of physiological measurements, which may be used to develop rules to assess probability or prognosis, or response to a stimulus; reproducibility of occurrence of a condition; and reproducibility of experimental results	Face Validity Reliability (Epidemiology) Reliability and Validity Reliability of Result Reliability of Results Reproducibility Of Result Reproducibility of Finding Reproducibility of Findings Test-Retest Reliability Validity (Epidemiology) Validity of Result Validity of Results

Appendix 3. Full search strategy and results by database

Feature	Report
Type of search	New
Database	Medline
Interface	Ovid
Search date	09/07/2022
Date filter	2002-2022

Nº	Search	Results
1	exp Self-Assessment/ or (((Assessment, Self or Self Assessment or Self-Criticism or Surveys) and Questionnaires) or Questionnaire or Questionnaires or Respondent or Respondents or Survey or Surveys).tw. or exp Disability Evaluation/	1105781
2	exp Young Adult/ or young adult.tw. or exp Adult/ or Adult.tw. or exp Middle Aged/ or Middle Aged.tw. or exp Aged/ or aged.tw. or exp "Aged, 80 and over"/ or (AGED, 80 and OVER).tw.	8396031
3	exp Patient Reported Outcome Measures/ or Outcome Assessment, Health Care/ or Patient Reported Outcome Measures.tw. or (Patient Reported Outcome or Patient Reported Outcome Measure or Patient Reported Outcomes or Patient-Reported Outcome or Patient-Reported Outcomes).tw.	103642
4	exp Psychometrics/ or psychometrics.tw. or Psychometric properties.tw. or exp "Reproducibility of Results"/ or (((Face Validity or Reliability or Reliability) and Validity) or Reliability of Result or Reliability of Results or Reproducibility of Result or Reproducibility of Finding or Reproducibility of Findings or Test-Retest Reliability or Validity or Validity of Result or Validity of Results).tw.	596210
5	#1 AND #2 AND #3 AND #4	3288
6	#5. Limited to last 10 years	2167
7	#6 Limited to validation study	694

Feature	Report
Type of search	New
Database	Embase
Interface	Embase
Search date	09/07/2022
Date filter	2002-2022

Nº	Search	Results
1	('assessment, self' OR (assessment, AND ('self'/exp OR self)) OR 'self assessment'/exp OR 'self assessment' OR (('self'/exp OR self) AND ('assessment'/exp OR assessment)) OR 'self criticism'/exp OR 'self criticism' OR 'surveys'/exp OR surveys) AND ('questionnaires'/exp OR questionnaires) OR 'questionnaire'/exp OR questionnaire OR 'questionnaires'/exp OR questionnaires OR respondent OR respondents OR 'survey'/exp OR survey OR 'surveys'/exp OR surveys OR 'disability evaluation'/exp OR 'disability evaluation' OR (('disability'/exp OR disability) AND ('evaluation'/exp OR evaluation))	2,614,378
2	(young AND adult OR adult OR middle) AND aged OR aged OR 'aged, 80 and over'	5,317,748
3	((((((patient AND reported AND outcome AND measures OR outcome) AND assessment, AND health AND care OR patient) AND reported AND outcome AND measures OR patient) AND reported AND outcome OR patient) AND reported AND outcome AND measure OR patient) AND reported AND outcomes OR 'patient reported') AND outcome OR 'patient reported') AND outcomes	123,694
4	((((((((((psychometrics OR psychometric) AND properties OR 'reproducibility of results' OR face) AND validity OR reliability) AND validity OR reliability) AND of AND result OR reliability) AND of AND results OR reproducibility) AND of AND result OR reproducibility) AND of AND finding OR reproducibility) AND of AND findings OR 'test retest') AND reliability OR validity) AND of AND result OR validity)	271,913
5	#1 AND #2 AND #3 AND #4	1,132
6	#5. Limited to last 10 years	974
7	#6 Limited to validation study	655

Feature	Report
Type of search	New
Database	Lilacs
Interface	Lilacs
Search date	09/07/2022
Date filter	2002-2022

Nº	Search	Results
1	Self-Assessment OR Surveys and Questionnaires OR Questionnaire OR Disability Evaluation	9,931
2	Young Adult OR Adult OR Middle Aged OR Aged OR "Aged, 80 and over"	35,317
3	Patient Reported Outcome Measures	270
4	Psychometric* OR Psychometric properties OR "Reproducibility of Results"	28,290
5	#1 AND #2 AND #3 AND #4	259
6	#5. Limited to last 10 years	224
7	#6 Limited to estudios pronóstico AND estudio observacional AND estudio diagnóstico AND estudio evaluación AND estudio tamizaje	21

Feature	Report
Type of search	New
Database	Scopus
Interface	Elsevier
Search date	09/07/2022
Date filter	2002-2022

Nº	Search	Results
1	(self-assessment OR surveys AND questionnaires OR disability AND evaluation)	76,949
2	(young AND adult OR adult* OR middle AND aged OR aged OR aged, 80 AND over)	190,523
3	(patient AND reported AND outcome AND measures)	73,223
4	("psychometrics" OR reproducibility AND of AND results)	490,102
5	#1 AND #2 AND #3 AND #4	66
6	#5. Limited to last 10 years	60
7	#6 Limited to "Article" AND "Article final"	54

Feature	Report
Type of search	New
Database	Rehabilitation Reference Center
Interface	EBSCO
Search date	14/07/2022
Date filter	2002-2022

Nº	Search	Results
1	Self-Assessment OR Assessment, Self OR Self Assessment OR Self-Criticism OR Surveys and Questionnaires OR Questionnaire OR Questionnaires OR Respondent OR Respondents OR Survey OR Surveys OR Disability Evaluation	1,197
2	Young adult OR Adult OR Adults OR Middle Aged OR Aged OR elderly OR aged, 80 and over	3,570
3	Patient Reported Outcome Measures (PROM) OR Patient Reported Outcome OR Patient Reported Outcome Measure OR Patient Reported Outcomes OR Patient-Reported Outcome OR Patient-Reported Outcomes	1,938
4	"psychometrics" OR Psychometric properties OR Reproducibility of Results OR Face Validity OR Reliability OR Reliability and Validity OR Reliability of Result OR Reliability of Results OR Reproducibility Of Result OR Reproducibility of Finding OR Reproducibility of Findings OR Test-Retest Reliability OR Validity OR Validity of Result OR Validity of Results	492
5	#1 AND #2 AND #3 AND #4	279
6	#5. Limited to last 10 years	274
7	#6 Limited to Limited to Revisiones clínicas	255

Feature	Report
Type of search	New
Database	Google Scholar
Interface	Google Scholar
Search date	10/07/2022
Date filter	2002-2022

Nº	Search	Results
1	Self-Assessment OR Surveys and Questionnaires OR Disability Evaluation	2,400,000
2	Young adult OR Adult* OR middle aged OR aged OR aged, 80 and over	2,860,000
3	Patient Reported Outcome Measures	5,860,000
4	"psychometrics" OR Reproducibility of Results	1,880,000
5	#1 AND #2 AND #3 AND #4	13,500
6	#5. Limited to last 10 years	7,960
7	#6 Limited to "validation Study" the first 200 records were taken.	1,110

Feature	Report
Tipo de búsqueda	New
Bases de datos	BVS
BVS	BVS
Fecha de búsqueda	15/07/2022
Rango de fecha de búsqueda	2002-2022

Nº	Search	Results
1	Self-Assessment OR Surveys and Questionnaires OR Questionnaire OR Disability Evaluation	270,816
2	Young Adult OR Adult OR Middle Aged OR Aged OR "Aged, 80 and over"	1,512,238
3	Patient Reported Outcome Measures	86,887
4	Psychometric* OR Psychometric properties OR "Reproducibility of Results"	4,047,210
5	#1 AND #2 AND #3 AND #4	259
6	#5. Limited to last 10 years	224
7	#6 Limited to "validation Study"	119

Appendix 4. Assessing risk of bias in a study on structural validity according to COSMIN

Structural Validity				
Instrument	For CTT: Was exploratory confirmatory factor analysis performed?	For IRT/Rasch: does the chosen model fit to the research question?	Was the sample size included in the analysis adequate?	Were there any other important flaw in the design or statistical methods of the study?
WHODAS II (Magistrale, et. al, 2015)	N. A	Adequate The subscale analysis, there was only a good fit in 4/7	Adequate >100 participants	No other important Methodologic flaws
WHODAS II (Snell, et al. 2017)	Adequate	N.A.	Doubtful Five times the number of items, but < than 100 participants	No other important Methodologic flaws
IMPACT-S (Schweinsberg, 2015)	Adequate	N.A.	Inadequate < 5 times the number of items	No other important Methodologic flaws

Internal Consistency				
Instrument	Was an internal consistency statistic calculated for each unidimensional scale or subscale separately?	For continuous scores: Was Cronbach´s alpha or omega calculated?	For dichotomous scores: Was Cronbach´s alpha or KR-20 calculated?	For IRT-based scores: Was standard error of the theta $(SE(\theta))$ reliability coefficient of estimated latent trait value (index of subject or item) separation calculated.
WHODAS II (Magistrale, et al. 2015)	Very good	Very good	N. A	N. A
WHODAS II (Snell, et al. 2017)	Inadequate Internal consistency was not reported for each factor. It was reported globally.	Very good	N.A.	N.A.
IMPACT-S (Schweinsberg, 2015)	Very good	Very good	N.A.	N.A.

Criterion Validity		
Instruments	For continuous scores: Were correlations, or the area under the receiver operating curve calculated?	For dichotomous scores: Were sensitivity and specificity determined?	Were there any other important flaws in the design or statistical methods of the study?
IMPACT-S (Schweinsberg, 2015)	Very good	N.A.	No other important Methodologic flaws

Reliability						
Instruments	Were patients stable in the interim period on the construct to be measured?	Was the time interval appropriate?	Were the test conditions similar for the measurements? E.g. Type of administration, environment, instructions	For continuous scores: Was an Intraclass Correlation Coefficient (ICC) calculated?	For dichotomous/ nominal/ ordinal scores: Was Kappa calculated?	For ordinal scores: Was the weighting scheme described? E.g. linear, quadractic
IMPACT-S (Schweinsberg, 2015)	Very good	Time Interval appropriate	Doubtful. Unclear if test conditions were similar	Very good. ICC calculated and model or formula of the ICC is described.	Very good	N.A.

NOTE: CTT= Classical Test Theory; **IRT=** Items Response Theory.

Hypotheses testing for construct validity			
Comparison with Other Outcome Measurement Instruments (Convergent Validity)			
Instruments	Is it clear what the comparator instrument (s) measure (s)?	Were the measurement properties of the comparator instrument(s) sufficient?	Were design and statistical methods adequate for the hypotheses to be tested?
WHODAS II (Magistrale, et al. 2015)	Very good	Very good	Very good
WHODAS II (Snell, et al. 2017)	N.A.	N.A.	N.A.
IMPACT-S (Schweinsberg, 2015)	Very good	Very good	Very good

Hypotheses Testing for Construct Validity		
Comparison Between Subgroups (Discriminative or Known-Groups Validity?		
Instruments	Was an adequate description provided of important characteristics of the subgroups?	Were design and statistical methods adequate for the hypotheses to be tested?
WHODAS II (Magistrale, et al. 2015)	N.A.	N.A.
WHODAS II (Snell, et al. 2017)	Very good	Very good
IMPACT-S (Schweinsberg, 2015)	N.A.	N.A.

Appendix 5 Generic hypotheses to evaluate construct validity and responsiveness[18]

1	Correlations with (changes) instruments measuring similar constructs should be ≥ 0.50
2	Correlations with (changes in) instruments measuring related, but dissimilar constructs should be lower i.e. $0.30 - 0.50$
3	Correlations with (changes in) instruments measuring unrelated constructs should be < 0.30
4	Correlations defined under 1, 2, and 3 should differ by a minimum of 0.10
5	Meaningful changes between relevant (sub) groups (e.g. patients with expected high vs low levels of the construct of interest)
6	For responsiveness, AUC should be ≥ 0.70

NOTE: AUC (Area Under the Curve)

Appendix 6. PROSPERO Protocol

A systematic review of the psychometric properties of disability instruments.
To enable PROSPERO to focus on COVID-19 submissions, this registration record has undergone basic automated checks for eligibility and is published exactly as submitted. PROSPERO has never provided peer review, and usual checking by the PROSPERO team does not endorse content. Therefore, automatically published records should be treated as any other PROSPERO registration. Further detail is provided here.

Citation
Camila Rodríguez Guevara, Diana Isabel Muñoz Rodríguez, Gino Montenegro Martínez, Pablo Esteban Roa Urrutia. Asystematic review of the psychometric properties of disability instruments. PROSPERO 2022 CRD42022348222 Available from: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42022348222

Review question
What are the instruments that, according to their psychometric properties, are recommended for measuring disability andfunctioning in people over 18 years of age?

Searches
The search will be carried out in databases and other sources of information. The databases will be Embase, MEDLINE (Ovid), Scopus, LILACS, Rehabilitation Reference Center (EBSCO), and BVS. The other sources of information will beGoogle Scholar and reverse snowball search or backward reference. The search strategy will build with four groups, the first of them related to assessment, evaluation, surveys, and questionnaires. The second group is people over 18 years old and keywords related to the target population; the third group is about Patient Reported Outcome Measures (PROM); finally, the fourth group includes reproducibility of results, psychometrics, and psychometric properties. The Boolean operator inside every group will be "OR" and between-group "AND". Restriction of language in English, Spanish and Portuguese. The search on Google Scholar can be constructed with three combinations and the one that retrieves the mostrecords will be selected. The snowball search will be reviewing the bibliography of articles included in this Systematic Review. The period is 10 years.

Search strategy
https://www.crd.york.ac.uk/PROSPEROFILES/348222_STRATEGY_20220723.pdf

Types of study to be included
-Validation study Independent of properties measurement analyzed.

-Study population of people over 18 years old to analyze disability and functioning which could include body functionsand structures, activities and participation, and environmental factors.

Condition or domain being studied
The condition of health to be analyzed is disability and functioning according to WHO (World Health Organization) landmark which defines disability as an umbrella term for impairments, activity limitations, and participation restrictions. It denotes the negative aspects of the interaction between an individual (with a health condition) and that individual's contextual factors (environmental and personal factors)" and functioning "is an umbrella term for body functions, body structures, activities, and participation. It denotes the positive aspects of the interaction between an individual (with a health condition) and that individual's contextual factors (environmental and personal factors)". (WHO, 2001). To determine disability and functioning status, the experts use different measurements such as Patient-Reported Outcome Measures (PROM), which are applied to know the patients' perceptions around disability and functioning status, and they are recommended by WHO and experts. (OMS, 2015).

Organización Mundial de la Salud. Clasificación Internacional del Funcionamiento, de la Discapacidad y de la Salud. 2011.

Organización Mundial de la Salud. Medición de la Salud y la Discapacidad. 2015.

Participants/population

-People over 18 years of age need an assessment of functional status and disability. These people can have different health conditions with diverse grades of severity.

-People over 18 years of age need an assessment of functional status and disability. These people without health conditions.

-At least 50% of the sample included in the validation of the instrument must represent the population of interest in thestudy.

Intervention(s), exposure(s)

The review will include profile, self-assessment, surveys, questionnaires, and disability and functioning assessments, overall.

Comparator(s)/control

No apply for this review.

Context

Types of study to be excluded:
-Studies different from validation studies such as clinical trials that use Patient Reported Outcome Measures (PROMS) to measure outcomes.
-Disability instruments different from Patient-Reported Outcome Measures such as biomechanical, muscle strength, joint range, and other functions evaluated by experts.

Main outcome(s)

For this kind of review, the main outcomes are related to measurement properties such as content validity, internal structure, and remaining measurement properties (reliability, error measurement, criterion validity, hypotheses testing for construct validity, and responsiveness). This taxonomy is proposed by COSMIN (Consensus-based selection of healthMeasurement Instruments).

Measures of effect

It depends on measurment properties.

Additional outcome(s)

See primary outcomes above.

Measures of effect

See primary outcomes above.

Data extraction (selection and coding)

Data extraction will have author and year, name of instruments, population characteristics, results on the measurement properties, additionally, interpretability (for instance, application time), and feasibility for each Patient Reported Outcome Measures (PROM) related disability and functioning assessment. On another hand, the researchers will showthe evaluation of the measurement properties depends on quality; for instance, either sufficient, insufficient, or indeterminate. Finally, the summary table with level evidence according to GRADE.

Risk of bias (quality) assessment

The risk of bias assessment will follow the COSMIN methodology, which included:

1. Bias checklist. To know which measurement properties are assessed in every article.

2. To assess the methodological quality of studies according to each measurement property. It can be scored as poor, fair, good, or excellent, which indicates the quality of every measurement.

Strategy for data synthesis

It depends on the possibility to synthesize quantitatively the results. If it is the case, it will make through meta-analysis for each measurement properties calculating weight means according to the number of participants. Additionally, it will have confidence intervals and use the random effects model. The heterogeneity will assess with the Higgins test.

Analysis of subgroups or subsets

First it all, The researcher will consider the importance to develop a subgroup analysis according to the results. If it is the case, the analysis will be for age groups and type of diagnosis with a meta-regression.

Contact details for further information

Camila Rodríguez Guevara
camilarodriguevara1@gmail.com

Organisational affiliation of the review

Universidad CES https://www.ces.edu.co/

Review team members and their organisational affiliations

Mrs. Camila Rodríguez Guevara. Estudiante doctorado en Epidemiología y Bioestadística. Universidad CES. Escuela de Graduados.
Dr. Diana Isabel Muñoz Rodríguez. Doctora en Epidemiología y Bioestadística. Facultad de Fisioterapia. Universidad CES. Coordinadora Investigación: Grupo de Investigación Movimiento y Salud. Medellín. Colombia.
Dr. Gino Montenegro Martínez. Doctor en Salud Pública. Coordinador Doctorado en Salud Pública. Escuela de Graduados, Universidad CES. Grupo de Investigación: Observatorio de la Salud Pública. Medellín. Colombia.
Mr. Pablo Esteban Roa Urrutia. Msc Epidemiología. Referente y coordinador equipo COVID-19. Secretaría de Salud Pública. Cali. Colombia.

Type and method of review

Systematic review, other.

Country

Colombia.

Published protocol

https://www.crd.york.ac.uk/PROSPEROFILES/348222_PROTOCOL_20220801.pdf

Stage of review [1 change]

Review completed, not published.

Subject index terms status

Subject indexing assigned by CRD.

Subject index terms

Humans; Psychometrics; Reproducibility of Results; Surveys and Questionnaires.

<div align="center">

Date of registration in PROSPERO

</div>

2 August 2022.

<div align="center">

Date of first submission

</div>

01 August 2022.

<div align="center">

Details of any existing review of the same topic by the same authors

</div>

There is not another review of the same topic by the same authors.

Stage of review at time of this submission [1 change]

Stage	Started	Completed
Preliminary searches	Yes	Yes
Piloting of the study selection process	Yes	Yes
Formal screening of search results against eligibility criteria	No	No
Data extraction	No	No
Risk of bias (quality) assessment	No	No
Data analysis	No	No

Revision note

The record owner confirms that the information they have supplied for this submission is accurate and complete and they understand that deliberate provision of inaccurate information or omission of data may be construed as scientific misconduct.

The record owner confirms that they will update the status of the review when it is completed and will add publication details in due course.

Versions

12 August 2022
12 August 2022
28 June 2022