ACRM
AMERICAN CONGRESS OF REHABILITATION MEDICINE

Archives of Physical Medicine and Rehabilitation

journal homepage: www.archives-pmr.org

Archives of Physical Medicine and Rehabilitation 2023;104: 656–72

REVIEW ARTICLE (META-ANALYSIS)

Measurement Properties of Clinical Instruments for Assessing Manual Wheelchair Mobility in Individuals With Spinal Cord Injury: Systematic Review

Gabriel Ribeiro de Freitas, PT, MSc,[a,b] Libak Abou, PT, PhD,[c] Aline de Lima, PT, MSc,[b] Laura A. Rice, PT, PhD,[d,e] Jocemar Ilha, PT, PhD[a,b]

From the [a]Neuroscience Graduate Program, Universidade Federal de Santa Catarina (UFSC), Florianópolis, Brazil; [b]Spinal Cord Injury Rehabilitation Research Group (SCIR-group), Department of Physiotherapy, College of Health and Sport Science, Universidade do Estado de Santa Catarina (UDESC), Florianópolis, Brazil; [c]Department of Physical Medicine and Rehabilitation, Michigan Medicine, University of Michigan, Ann Arbor, MI; [d]Department of Kinesiology and Community Health, College of Applied Health Sciences, the University of Illinois at Urbana-Champaign, Urbana, IL; and [e]Center on Health, Aging, and Disability, College of Applied Health Sciences, the University of Illinois at Urbana-Champaign, Urbana, IL.

Abstract

Objective: To evaluate the measurement properties of clinical instruments used to assess manual wheelchair mobility in individuals with spinal cord injury (SCI).

Data Sources: This systematic review was conducted according to the Consensus-Based Standards for the Selection of Health Measurement Instruments guidance and Preferred Reporting Items for Systematic Reviews and Meta-Analyses statement. The search was conducted up to December 2021 on MEDLINE/PubMed, Cochrane Central Register of Controlled Trials, Web of Science, *Biblioteca Regional de Medicina*, and Cumulative Index to Nursing and Allied Health databases without time restriction.

Study Selection: Peer-reviewed original research articles that examined any clinical wheelchair mobility and/or skill assessment instrument among adults with SCI and reported data on at least one measurement property or described the development procedure were evaluated independently by two reviewers.

Data Extraction: Data were independently extracted according to Consensus-Based Standards for the Selection of Health Measurement Instruments methodology. Measurement property results from each study were independently rated by two reviewers as sufficient, insufficient, indeterminate, or inconsistent. The evidence for each measurement property was rated as high, moderate, low, or very low (Grading of Recommendations, Assessment, Development, and Evaluation). Recommendations for highly-rated instruments were performed.

Data Synthesis: Twenty-nine studies with 21 instruments were identified. The methodological quality of studies ranged from insufficient to sufficient, and the quality of evidence ranged from very low to high. Six instruments reported content validity. Reliability and construct validity were the most studied measurement properties. Structural validity and invariance for cross-cultural measurement were not reported. The highly rated instruments were the Wheelchair Outcome Measure and Wheelchair Skills Test Questionnaire.

Conclusions: Although numerous instruments for assessing wheelchair mobility and/or skills among individuals with SCI were identified, not many measurement properties have been sufficiently established. The Wheelchair Outcome Measure and Wheelchair Skills Test Questionnaire show the current best potential to be recommended for clinical and research use. Further studies are needed to strengthen or change these recommendations.

Archives of Physical Medicine and Rehabilitation 2023;104:656–72

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Brazil, Finance Code 001 and Fundação de Amparo à Pesquisa e Inovação do Estado de Santa Catarina (FAPESC), Brazil, TO 2019TR767.

Systematic Review Registration No. (PROSPERO): CRD42019138301.

Disclosures: none.

Individuals living with spinal cord injury (SCI) may present several functional limitations, including significant mobility limitations.[1,2] Approximately 80% of individuals living with SCI use a wheelchair as their primary form of mobility to perform

0003-9993/$36 - see front matter © 2022 by the American Congress of Rehabilitation Medicine.
https://doi.org/10.1016/j.apmr.2022.10.002

functional tasks.[3] Essential wheelchair skills are related to proper propulsion mechanics, transfers between the wheelchair and other surfaces, maintenance of the wheelchair, and management of environmental barriers, such as ramps, curbs, and stairs.[4,5]

Therefore, the ability to perform mobility activities in a wheelchair is essential for the individual's independence in daily life activities, providing security during their movement and increasing social participation and quality of life.[6-8] Other essential aspects are related to physical capacity, return to work activity, personal satisfaction, and greater mobility capacity, which allow the individual to remain active and reduce sedentary lifestyle and deconditioning.[6,9]

Using standardized assessment instruments to assess wheeled mobility skills is essential to classify individuals' skill levels and guide the training of appropriate wheelchair skills.[8,10] Many different instruments are available for use in clinical practice.[10-12] However, most of them were not developed specifically for individuals living with SCI (target population).[10] Furthermore, there is no consensus in the literature about which instruments are the most suitable and reliable for evaluating mobility construct in individuals with SCI who use a wheelchair for daily activities.[10,11] Additionally, the process leading to the development of several of these outcome measures is not standardized, presenting limitations in their methodologies, which hinders the interpretation of the measurement properties.[10,11]

To standardize the assessment of the measurement properties, the Consensus-Based Standards for the Selection of Health Measurement Instruments (COSMIN) initiative can guide the development of outcome measures.[13] It can also be used to conduct a methodological quality assessment of the measurement properties of instruments.[14,15] Analysis with the COSMIN guidelines can help in the decision-making process to select the most suitable clinical tools for use in research and clinical settings, allowing for determining recommendation levels for diagnostic tests.[14,16] The COSMIN developed a methodology guideline for developing systematic reviews of patient-reported outcome measures (PROMs),[14,16] which uses a systematic methodology with defined criteria to standardize such type of study.

To our knowledge, 3 previous reviews on a similar subject were published more than 10 years ago.[10,17] The studies aimed to identify and analyze wheelchair mobility abilities and how they could be measured. However, they did not use the COSMIN taxonomy and systematic review methodology for the measurement property analysis of each instrument. In addition, only 1 of the previous reviews analyzed wheelchair skills tests specifically among individuals with SCI who use a manual wheelchair for their mobility activities.[10] Moreover, these previous reviews have shown significant inconsistencies among the instruments available more than 1 decade ago. The studies have also failed to make any recommendation on the most suitable instrument to assess wheelchair mobility skills among individuals with SCI.[10,17,]

Therefore, this study aims to systematically review the instruments available in the literature to assess wheelchair mobility among individuals with SCI. In addition, this review seeks to recommend outcome measurement instruments to evaluate wheelchair mobility based on their measurement properties.

List of abbreviations:

AIS	**American Spinal Injury Association Impairment Scale**
COSMIN	**Consensus-Based Standards for the Selection of Health Measurement Instruments**
5AML-FIM	**5 Additional Mobility and Locomotor in FIM**
GRADE	**Grading of Recommendations, Assessment, Development, and Evaluation**
PROM	**patient-reported outcome measure**
QEWS	**Queensland Evaluation of Wheelchair Skills**
SCI	**spinal cord injury**
SEWM	**Self-efficacy in Wheeled Mobility Scale**
TOWM	**Test of Wheeled Mobility**
WC-PFP	**Wheelchair Physical Functional Performance**
WhOM	**Wheelchair Outcome Measure**
WPT	**Wheelchair Propulsion Test**
WST	**Wheelchair Skills Test**
WST-Q	**Wheelchair Skills Test Questionnaire**
WUFA	**Wheelchair Users Functional Assessment**

Methods

This systematic review was conducted according to the COSMIN methodology and systematic review guideline[14,16] and reported using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses statement.[18,19] The research protocol was developed and registered in the Prospective International Registry of Systematic Reviews (http://www.crd.york.ac.uk/PROSPERO/; registration no.: CRD42019138301). For the purpose of this review, instruments for assessing manual wheelchair mobility were defined as any clinical scale, test, or questionnaire used to assess mobility-related activities to move by changing body position or location or by transferring from one place to another using a manual wheelchair.

Data sources and searches

A literature search was conducted in December 2021 using the following electronic databases: MEDLINE/PubMed, Cochrane Central Register of Controlled Trials, Web of Science, Biblioteca Regional de Medicina, and Cumulative Index to Nursing and Allied Health. The general search strategy consisted of 3 search groups related to (1) SCI, (2) wheelchairs, and (3) outcome assessment. Each group was composed of the Medical Subject Headings search terms and/or synonyms and variations, as shown in appendix 1. The complete search strategy was adapted for each electronic database. No date or language restrictions were applied during the search. Finally, a forward and backward investigation of the eligible studies was carried out by checking their reference lists for additional studies.

Study selection

The initial screening of titles was performed to avoid duplicate references and exclude apparent nonfitting studies. Then, the remaining full-text articles deemed potentially relevant were reviewed for eligibility. In both stages, 2 reviewers performed screening independently (G.R.F., L.A.). Any discrepancies in the agreement were solved with a third reviewer (J.I.) until a consensus was reached.

Articles were deemed eligible when the study (1) reported that the sample contained adults with SCI who are manual wheelchair users, (2) discussed any clinical wheelchair mobility or skill assessment instrument, and (3) reported data on at least 1 measurement property or described the development procedures of a clinical instrument evaluating wheelchair mobility or skill. Articles

were excluded if they were systematic reviews, study proposals or protocols, case studies or case series, books or book chapters, poster abstracts, conference abstracts, correspondence, or commentaries.

Data extraction

One reviewer (G.R.F.) extracted all necessary data using a standard extraction form, checked by a second reviewer (L.A.). The initial data extracted from the included articles were (1) author, (2) country/language, (3) aim of the study, (4) aspects of the study sample, (5) measurement properties studied, and (6) instrument characteristics (number of items, type, and administration).

Assessment of the methodological quality of the included studies

According to the recommendations of the COSMIN checklist risk of bias, 2 reviewers (A.L., G.R.F.) independently assessed the methodological quality of the included studies.[14,16] The COSMIN checklist is a tool that contains 10 boxes to evaluate the methodological quality of each study on measurement properties and 2 boxes to evaluate general requirements.[13,15]

The COSMIN guidance for systematic reviews of PROMs recommends assessing measurement properties in the following order: (1) content validity and (2) internal structure (test-retest reliability, measurement error, criterion validity, construct validity, responsiveness).[16] For each measurement property, 3 phases are included. Initially, the risk of bias in every single study is assessed. Afterward, the results of every single study are rated against the criteria for sufficient measurement properties. Finally, the results from all studies are summarized, and the quality of evidence is graded for each measurement property.[14,16]

Assessment of the outcome measurement properties

In the first phase, the risk of bias was evaluated. Each measurement property was independently classified as very good (V), adequate (A), doubtful (D), or inadequate (I) by 2 reviewers (A.L., G.R.F.). To rate the methodological quality, the lowest rating of any item in a COSMIN evaluation box was taken (worse score counts method) according to the COSMIN guidelines, and in case of disagreement, a third evaluator was consulted (J.I.).[14,16]

Next, the criteria for sufficient measurement properties of each article were analyzed by the same 2 independent reviewers. In this phase, every measurement property of each article is rated according to the consensus-based criteria proposed by COSMIN.[14,16,20] Based on previous studies, the review team formulated a set of hypotheses (fig 1 - box) to evaluate construct validity (hypothesis testing or known-groups validity) and responsiveness.[14,20,21]

After that, an overall rating for each measurement property was used, combining all studies that evaluated a specific instrument. Finally, the instrument's measurement property was classified as sufficient, insufficient, inconsistent, or indeterminate.[14,16] Construct validity and responsiveness were considered sufficient if >75% of the hypotheses were met, insufficient if <75% of the hypotheses were not met, or inconsistent if different studies showed divergent findings, that is, if they did not display the same results (eg, all sufficient, insufficient, inconsistent findings).[14,16]

Best evidence synthesis

The quality of the evidence was graded for every measurement property and each instrument separately according to the Grading of Recommendations, Assessment, Development, and Evaluation (GRADE) approach adapted for this type of review (4-point score: high, moderate, low, or very low).[22–24] If the measurement property studied did not achieve the criteria, the score was downgraded, and confidence in the instrument was reduced.[23] We determined the requirements for GRADE indirectness evaluation. GRADE was downgraded (decrease score) for a mixed population: (−1) if >75% with SCI and (−2) if the sample contained between 75%-50% of participants living with SCI.[24] The quality of evidence (GRADE: high, moderate, low, very low) was combined with study outcomes rating (COSMIN: positive, indeterminate, negative as proposed by the Cochrane Back Review Group).[25]

After the interpretation of the evidence synthesis, recommendations for highly-rated instruments were performed according to the COSMIN guideline into 3 categories: (A) instruments that have the potential to be recommended as the most suitable for the construct and population of interest (ie, instruments with evidence for sufficient content validity [any level] and at least low-quality evidence for sufficient internal consistency); (B) instruments that have the potential to be recommended, but more validation studies are necessary (ie, instruments categorized not in A or C); and (C) instruments that should not be recommended (ie, instruments with high-quality evidence for an insufficient measurement property).[14,16]

Results

Characteristics of included studies

In total, 2927 records were identified in the database searches, 59 full-text articles were retrieved, and 29 studies met the eligibility criteria (fig 2, appendix 2). Among the 29 included studies in this review, 21 instruments were identified, as shown in table 1 (17 tests and 4 questionnaires). The sample characteristics were heterogeneous among the studies. They included individuals with paraplegia and tetraplegia from American Spinal Injury Association) Impairment Scale (AIS) A-D classifications. Moreover, individuals with SCI from inpatient/outpatient settings and from subacute to chronic stages were included in the studies (tables 2-4). The overall methodological quality of the studies ranged from insufficient to sufficient (see tables 2-4), and the quality of evidence of the outcome measurement instruments ranged from very low to high (table 5).

Measurement property analysis

All 21 instruments included in this review analyzed at least 1 measurement property. Of these, only 6 instruments (28%) had the content validity, and 5 instruments (23%) had the internal consistency analyzed.

The content validity was analyzed (see table 2) in the Test of Wheeled Mobility (TOWM), Wheelchair Skills Test (WST), Wheelchair Skills Test Questionnaire (WST-Q), Wheelchair Outcome Measure (WhOM), Wheelchair Users Functional Assessment (WUFA), and Wheelchair Propulsion Test (WPT). Among

Measurement property	Criteria
Construct Validity	• Correlation with an instrument measuring the same construct must be ≥0.50 OR at least 75% of the results are in accordance with the hypotheses AND correlation with related constructs is higher than with unrelated constructs **(+)** • <0.50 and ≥0.30 with instruments measuring related but dissimilar constructs (e.g., quality of life, life satisfaction); and <0.30 with instruments measuring unrelated constructs **(+)** • Solely correlations determined with unrelated constructs OR no correlations reported **(?)** • Correlation with an instrument measuring the same construct is <0.50 OR <75% of the results are in accordance with the hypotheses AND correlation with related constructs is higher than with unrelated constructs **(-)**
Responsiveness	• Area under the curve (AUC) to discriminate between improved and not improved/deteriorated subjects had to be ≥0.70 **(+) OR;** • Effect sizes and standardized response means for improved subjects had to be ≥0.50 larger than those for not improved/deteriorated subjects **(+) OR;** • p<0.05 AND change ≥20% from baseline, **(+)** • (?) only p<0.05 value (-) criteria for (+) do not match.
Criterion Validity	• Convincing arguments that criterion validity standard is "gold" AND correlation OR ICC with criterion standard is >0.70; OR sensitivity/specificity or accuracy is 80%-100% **(+).** • No convincing arguments that criterion standard is "gold" OR no correlations have been calculated. Sensitivity and specificity are not displayed OR only the accuracy is displayed **(?)** • Correlation with criterion standard is <0.70, despite adequate design and methods OR sensitivity and specificity or accuracy is <70% **(-)**

Fig 1 Criteria for hypotheses testing as defined by the review team. Abbreviations: AUC, area under the curve; ICC, intraclass correlation coefficient; OR, odds ratio.

those instruments, the TOWM, WPT, and WUFA were rated as very low−quality evidence of indeterminate content validity; the WhOM and WST-Q were rated as low-quality evidence of sufficient content validity; and the WST was rated as moderate-quality evidence of indeterminate content validity (see table 5).

The internal consistency (see table 2) was assessed in 5 instruments, those being the Five Additional Mobility and Locomotor in FIM (5AML-FIM), Queensland Evaluation of Wheelchair Skills (QEWS), Self-efficacy in Wheeled Mobility Scale (SEWM), Wheelchair Physical Functional Performance (WC-PFP), and WUFA. Moreover, these instruments were rated as having very low− to low-quality evidence of indeterminate internal consistency (see table 5).

Reliability was evaluated in 16 instruments (76%) and was the measurement property most evaluated among the included studies (see table 2). The analysis indicates that only the WhOM and WST showed criteria to be rated as moderate- or high-quality evidence (respectively) of sufficient reliability (see table 5). Construct validity (hypotheses testing) was evaluated among 12 instruments (57%) (see table 3). Four instruments were rated as having high- or moderate-quality evidence of sufficient construct validity (5AML-FIM, QEWS, WhOM, WPT). The other instruments were poorly rated and obtained a rating of inconsistent, indeterminate, or insufficient construct validity (see table 5).

Criterion validity was assessed by 8 instruments (38%), see table 3. Among those instruments, 3 (QEWS, WPT, WST-Q) were rated as having very low− and low-quality evidence of sufficient criterion validity (see table 5). The remaining 5 instruments were rated as having low- to moderate-quality evidence for insufficient or indeterminate criterion validity.

The measurement error, interpretability, and responsiveness analysis are shown in table 4. Four instruments had the measurement error analyzed (Short Wheelie Test, TOWM, Timed Manual Wheelchair Slalom Test, WhOM). Among those, the WhOM was

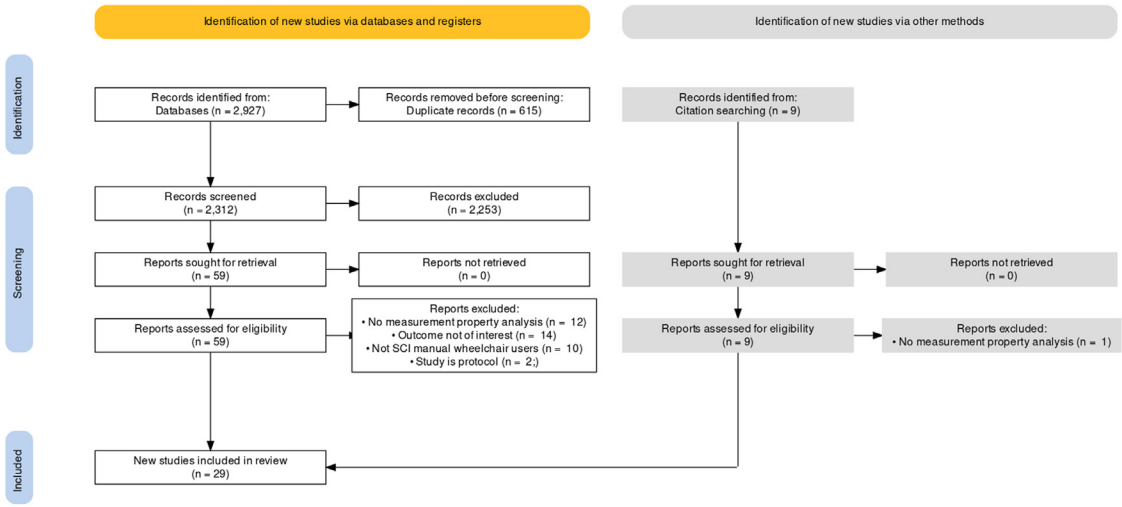

Fig 2 Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram.

Table 1 Included instruments and their characteristics

Instrument	Acronym/ Initials	No. of Items	Country (Language)	Mode of Administration	Reference
Adapted Manual Wheelchair Circuit	AMWC	14	US (English)/Brazil (Portuguese)	Test	Cowan et al[26]; Ribeiro Neto et al[27]
Adapted Manual Wheelchair Circuit+3	AMWC+3	17	Brazil (Portuguese)	Test	Ribeiro Neto et al[28]
Assessment tool for mobility in wheelchairs	ATMW	6	Australia (English)	Test	Harvey et al[29]
Comprehensive Product-Centered Approach to Assessing Functional Performance	PWFPA	14	Spain (Spanish)	Test	Gil-Agudo et al[30]
5 Additional Mobility and Locomotor in FIM	5AML-FIM	23	Australia (English)	Interview	Middleton et al[31]
Functional Tasks	FT	4	Canada (English)	Test	May et al[32]
Obstacle Course Assessment Wheelchair User Performance	OCAWUP	10	Canada (French)	Test	Routhier et al[33]
Performance-Based Wheelchair Propulsion Test	PWPT	4	Canada (English)	Test	Gagnon et al[34]
Queensland Evaluation of Wheelchair Skills	QEWS	5	Australia (English)	Test	Gollan et al[35]
Self-efficacy in Wheeled Mobility Scale	SEWM	10	Belgium (Dutch)	Interview	Fliess-Douer et al[36]
Short Wheelie Test	WT-S	8	Belgium (Dutch)	Test	Fliess-Douer et al[37]; Fliess-Douer et al[38]
9-Task Wheelchair Circuit	9-WC	9	Netherlands (Dutch)	Test	Kilkens et al[39]
Test of Wheeled Mobility	TOWM	30	Belgium (Dutch)	Test	Fliess-Douer et al[37]; Fliess-Douer et al[38]
Timed Manual Wheelchair Slalom Test	TMWST	1	Canada (English)	Test	Gagnon et al[40]
Wheelchair Circuit	WC	8	Canada (English)	Test	Kilkens et al
Wheelchair Outcome Measure	WhOM	Variable	Canada (English)/Iran (Farsi)	Interview	Mortenson et al[41]; Miller et al[42]; Alimohammad et al[43]
Wheelchair Physical Functional Performance	WC-PFP	12	US (English)	Test	Cress et al[44]
Wheelchair Propulsion Test	WPT	3	Canada (English)	Test	Askari et al[45]
Wheelchair Skill Test	WST	32	Canada (English)	Test	Kirby et al[46]; Kirby et al[47]; Kirby et al[48]; Lindquist et al[49]; Pradon et al[33]
Wheelchair Skill Test Questionnaire	WST-Q	32	Canada/US (English)	Interview (phone, postal questionnaire, online)	Mountain et al[50]; Kirby et al[46]; Passuni et al[51]
Wheelchair Users Functional Assessment	WUFA	13	US (English)	Test	Stanley et al[52]

Table 2 Measurement properties of the included instruments: internal consistency, reliability, and content validity

			Internal Consistency		Reliability		Content Validity					
							Relevance		Comprehensiveness		Comprehensibility	
Instrument	Reference	Sample	Method Quality	Result (Rating)	Method Quality	Result (Rating)	Method Quality	Result (Rating)	Method Quality	Result (Rating)	Method Quality	Result (Rating)
AMWC	Cowan et al[26]	50 Participants (100% SCI) Setting: outpatient Classification: paraplegia/tetraplegia AIS: A-D Chronicity: chronic			D	ICC=0.90 (+)						
ATMW	Harvey et al[29]	20 Participants (100% SCI) Setting: NR Classification: Paraplegia AIS: NR Severity: NR Chronicity: chronic			D	Cohen weighted κ=0.82-0.96 (+)						
PWFPA	Gil-Agudo et al[30]	6 Participants (100% SCI) Setting: NR Classification: paraplegia/tetraplegia (T1-T12) AIS: NR Chronicity: chronic			D	ICC=0.95-0.99 (−)						
5AML-FIM	Middleton et al[31]	43 Participants (100% SCI) Setting: NR Classification: paraplegia/tetraplegia AIS: NR Chronicity: NC	V	Cronbach α=0.89 (+)								
FT	May et al[32]	20 Participants (100% SCI) Setting: NR Classification: NR AIS: NR Chronicity: NR			V	ICC=0.99 (+)						
OCAWUP	Routhier et al[33]	17 Participants (35.3% SCI and 64.7% Other) Setting: NR Classification: NR AIS: NR Chronicity: chronic			A	Time (ICC=0.74 - 0.99) (+)/GSE (ICC>0.96) (+)						
PWPT	Gagnon et al[34]	14 Participants (100% SCI) Setting: inpatient Classification: tetraplegia and paraplegia (C6-L1) AIS: NR Chronicity: subacute			D	20-m propulsion test (ICC=0.981)/Slalom Test (ICC=0.978)/6-minute propulsion test (ICC=0.98)						
QEWS	Gollan et al[35]	20 participants (100% SCI) Setting: NR Classification: NR AIS: NR Chronicity: subacute to chronic	A	Cronbach α=0.86 (+)	A	ICC=0.98-1.00 (+)						

(*continued on next page*)

Table 2 (*Continued*)

Instrument	Reference	Sample	Internal Consistency		Reliability		Content Validity					
							Relevance		Comprehensiveness		Comprehensibility	
			Method Quality	Result (Rating)	Method Quality	Result (Rating)	Method Quality	Result (Rating)	Method Quality	Result (Rating)	Method Quality	Result (Rating)
SEWM	Fliess-Douer et al[36]	47 Participants (100% SCI) Setting: outpatient Classification: paraplegia AIS: NR Chronicity: chronic	V	Cronbach α=0.91/ r=0.90 (+)								
WT-S	Fliess-Douer et al., 2013[37]	30 Participants (100% SCI) Setting: outpatient Classification: NR AIS: NR Chronicity: chronic			A	ICC=0.94-0.99 (+)						
9-WC	Kilkens et al. 2002[39]	27 participants (100% SCI) Setting: Outpatient Classification: tetraplegia/paraplegia AIS: NR Chronicity: NR			A	ICC=0.76-0.98 (±)						
TOWM	Fliess-Douer et al[37]	30 Participants (100% SCI) Setting: outpatient Classification: NR AIS: NR Chronicity: chronic			A	ICC=0.91-0.99 (+)	D	?	D	?	D	?
TMWST	Gagnon et al[40]	15 Participants (100% SCI) Setting: NR Classification: tetraplegia/paraplegia AIS: A, B, or C Chronicity: subacute to chronic			D	eRliability coefficient φ=0.981 (?)						
WhOM							I		D	?	D	+
	Miller et al[42]	50 Participants (100% SCI) Setting: outpatient Classification: tetraplegia/paraplegia AIS: A, B, or C Chronicity: chronic			V	Sat (ICC=0.91)/ Imp × Sat (ICC=0.90) (+)						
	Alimohammad et al[43]	75 Participants (100% SCI) Setting: NR Classification: NR AIS: NR Chronicity: subacute to chronic			V	ICC=0.99 (+)						
WC-PFP	Cress et al[44]	18 Participants (44% SCI and 66% other) Setting: NR Classification: NR AIS: NR Chronicity: chronic	A	Cronbach α=0.93 (+)	A	ICC=0.96 (+)						
WPT	Askari et al[45]	58 Participants (8.6% SCI and 91.4% other) Setting: inpatient (96%)/outpatient (4%) Classification: NR AIS: NR Chronicity: acute and subacute			A	ICC=0.72-0.96 (+)	A	+	A	+	A	+

(*continued on next page*)

Table 2 (*Continued*)

			Internal Consistency		Reliability		Content Validity					
							Relevance		Comprehensiveness		Comprehensibility	
Instrument	Reference	Sample	Method Quality	Result (Rating)	Method Quality	Result (Rating)	Method Quality	Result (Rating)	Method Quality	Result (Rating)	Method Quality	Result (Rating)
WST	Kirby et al[47]	24 Participants (12.5% SCI and 87.5% other) Setting: inpatient/outpatient Classification: tetraplegia/paraplegia AIS: NR Chronicity: acute to chronic			A	ICC=0.65-0.96 (±)	D	?	I	?	D	?
	Kirby et al[48]	169 Participants (11% SCI and 89% other) Setting: inpatient/outpatient Classification: tetraplegia/paraplegia AIS: NR Chronicity: subacute to chronic			A	ICC=0.904-0.968 (±)	D	?	D	?	D	?
	Kirby et al[46]	117 Participants (100% SCI) Setting: outpatient Classification: tetraplegia/paraplegia AIS: NR Chronicity: chronic					D	?	D	?	I	?
	Lindquist et al[49]	11 Participants (81.8% SCI and 18.2% other) Setting: outpatient Classification: NR AIS: NR Chronicity: NC			A	ICC=0.85-0.90 (±)						
	Pradon et al[33]	40 Participants (100% SCI) Setting: NR Classification: paraplegia/tetraplegia AIS: NR Chronicity: chronic			A	ICC=0.84-0.94 (±)						
WST-Q	Mountain et al[50]	20 Participants (20% SCI and 80% other) Setting: inpatient Classification: NR AIS: NR Chronicity: subacute to chronic					I	+	I	+	I	+
	Kirby et al[46]	117 Participants (100% SCI) Setting: outpatient Classification: paraplegia/tetraplegia AIS: NR Chronicity: chronic					I	+	D	+	D	+
	Passuni et al[51]	11 Participants (100% SCI) Setting: NR Classification: paraplegia/tetraplegia (T12-C4) AIS: A-D Chronicity: NC			A	ICC=0.998 (+)						

(*continued on next page*)

Table 2 (*Continued*)

Instrument	Reference	Sample	Internal Consistency		Reliability		Content Validity					
							Relevance		Comprehensiveness		Comprehensibility	
			Method Quality	Result (Rating)	Method Quality	Result (Rating)	Method Quality	Result (Rating)	Method Quality	Result (Rating)	Method Quality	Result (Rating)
WUFA	Stanley et al[52]	5 Participants (60% SCI and 40% other) Setting: NR Classification: paraplegia/tetraplegia (T9 to T11) AIS: A and B Chronicity: chronic	V	Coefficient α=0.96 (+)	A	ICC=0.96 (+)	D	?	I	?	I	?

NOTE. Overall rating: −, insufficient; +, sufficient; ±, inconsistent; ?, indeterminate.

Abbreviations: A, adequate; AMWC, Adapted Manual Wheelchair Circuit; AMWC+3, Adapted Manual Wheelchair Circuit+3; ATMW, Assessment Tool for Mobility in Wheelchairs; D, doubtful; FT, Functional Tasks; I, inadequate; ICC, intraclass correlation coefficient; Imp, importance; NC, not clear; 9-WC, 9-Task Wheelchair Circuit; NR, not reported; OCAWUP, Obstacle Course Assessment Wheelchair User Performance; PWFPA, Comprehensive Product-Centered Approach to Assessing Functional Performance; PWPT, Performance-Based Wheelchair Propulsion Test; Sat, satisfaction; TMWST, Timed Manual Wheelchair Slalom Test; V, very good; WC, Wheelchair Circuit; WT-S, Short Wheelie Test.

rated as having moderate-quality evidence of sufficient measurement error, and the other 3 instruments were rated as having very low− and low-quality evidence of sufficient measurement error (see table 5).

Responsiveness was assessed by 3 instruments (14%) (see table 4). The Performance-Based Wheelchair Propulsion Test, Wheelchair Cicuit, and 5AML-FIM were rated as having very low − to low-quality evidence of sufficient responsiveness (see table 5). Additionally, the interpretability for 5 instruments (Functional Tasks, Short Wheelie Test, TOWM, Timed Manual Wheelchair Slalom Test) was reported in (see table 4). The remaining measurement properties, the structural validity, and the invariance for cross-cultural measurement were not evaluated by any study.

Best evidence synthesis and recommendations

The best evidence synthesis of each measurement property by the instruments, as reported in table 5, was summarized as the COSMIN recommendations in (see table 6). All the instruments included in this review were classified as category B (potential for use, but more studies are needed). Additionally, because the WhOM and WST-Q were the highest rated instruments for at least sufficient content validity, they can present the current best potential to be recommended for use.

Discussion

This systematic review synthesized the methodological assessment of the measurement properties for 21 instruments used to evaluate wheelchair mobility skills among individuals with SCI who use manual wheelchairs. In general, assessment instruments with insufficient measurement properties had been used to assess wheelchair mobility among individuals with SCI in clinical practice and research settings. Despite this, most of the assessment instruments demonstrated appropriate validation and accuracy according to the COSMIN criteria and were recommended in category B.[14,16] These instruments have the potential to be recommended for use, but more validation studies are necessary for the specific population, that is, individuals with SCI who use a manual wheelchair.[14]

Content validity is the first measurement property considered when choosing an instrument.[13] This is because if the proposed construct by an outcome measurement instrument is unclear, the other measuring properties will not be suitable for that outcome within the proposed construct.[13,15] Our systematic review uncovered only 6 instruments that had the content validity evaluated. Three instruments were rated as very low−quality evidence of indeterminate content validity (TOWM, WPT, WUFA), 1 instrument was rated as moderate-quality evidence of indeterminate content validity (WST), and 2 instruments were rated as having low-quality evidence of sufficient content validity (WhOM, WST-Q).

The second most relevant measurement property for instrument selection is internal consistency.[13,15] Measurement of both content validity and internal consistency is related to the degree of interrelationship between the instrument's items and its scores, reflecting the dimensionality of the construct to be measured.[14,15] Internal consistency was evaluated in 5 instruments rated as having very low− to low-quality evidence of indeterminate internal consistency (5AML-FIM, SEWM, QEWS, WC-PFP, WUFA). Only 1 instrument (WUFA) had both content validity and internal

Table 3 Measurement properties of the included instruments: validity

			Validity					
			Hypothesis Testing				Criterion	
			Convergent and Divergent		Known Groups			
Instrument	Reference	Sample	Method Quality	Result (Rating)	Method Quality	Result (Rating)	Method Quality	Result (Rating)
AMWC	Cowan et al[26]	50 Participants (100% SCI) Setting: outpatient Classification: paraplegia/tetraplegia AIS: A-D Chronicity: chronic			V	2 of 3 hypotheses confirmed	D	Discriminate between PP and TP subgroups sum ability (PP, 12.91.2; TP, 9.82.8; $P<.001$) (-)
	Ribeiro Neto et al[27]	66 Participants (100% SCI) Setting: NR Classification: paraplegia/tetraplegia (C5-L3) AIS: NR Chronicity: chronic			V	4 of 5 hypotheses were confirmed		
AMWC+3	Ribeiro Neto et al[28]	126 Participants (100% SCI) Setting: NR Classification: paraplegia AIS: A or B Chronicity: chronic					A	Discriminate all paraplegia subgroups (Wilks λ=0.86) (+)
5AML-FIM	Middleton et al[31]	43 Participants (100% SCI) Setting: NR Classification: paraplegia/tetraplegia AIS: NR Chronicity: NC			V	Hypothesis confirmed (+)	V	Discriminate between TP and PP groups ($P<.0001$) (—)
OCAWUP	Routhier et al[33]	17 Participants (35.3% SCI and 64.7% other) Setting: NR Classification: NR AIS: NR Chronicity: chronic	V	GSE and FIM partial score (r=0.84)	A	Hypothesis confirmed		
PWPT	Gagnon et al[34]	14 Participants (100% SCI) Setting: inpatient Classification: tetraplegia and paraplegia (C6- L1) AIS: NR Chronicity: subacute	I	Slalom × 20-m propulsion at maximum speed (admission r=0.93) and (discharge r=0.92)				
QEWS	Gollan et al[35]	20 Participants (100% SCI) Setting: NR Classification: NR AIS: NR Chronicity: subacute to chronic	V	QEWS × Concurrent Global Rating of Wheelchair Skills (r=0.93)			V	Correlation between the QEWS total score and the concurrent global rating of wheelchair skills (r^2=0.93) (+)
SEWM	Fliess-Douer et al[36]	47 Participants (100% SCI) Setting: outpatient Classification: paraplegia AIS: NR Chronicity: chronic	V	(Spearman correlation) SEWM × GSES =0.50; SEWM × ESES =0.60			V	Correlations between scales (SEWM-ESES: 0.60) and (SEWM-GSES: 0.50) ($P<.05$) (—)
TOWM	Fliess-Douer et al[38]	30 Participants (100% SCI) Setting: outpatient Classification: NR AIS: NR Chronicity: chronic	V	TOWM × WT-S =0.84; $P<.001$	V	TOWM r=0.45		
WC	Kilkens et al[53]	74 Participants (100% SCI) Setting: NR Classification: tetraplegia/paraplegia AIS: NR Chronicity: NR			V	4 of 5 hypotheses confirmed		

(*continued on next page*)

Table 3 (*Continued*)

			Validity					
			Hypothesis Testing				Criterion	
			Convergent and Divergent		Known Groups			
Instrument	Reference	Sample	Method Quality	Result (Rating)	Method Quality	Result (Rating)	Method Quality	Result (Rating)
WhOM	Miller et al[42]	50 Participants (100% SCI) Setting: outpatient Classification: tetraplegia/paraplegia AIS: A, B, or C Chronicity: chronic	V	K=0.71 (+)				
	Alimohammad et al[43]	75 Participants (100% SCI) Setting: NR Classification: NR AIS: NR Chronicity: subacute to chronic	V	Total Imp/Sat × SCIM-III (r=0.38)	V	Hypothesis confirmed (+)		
WC-PFP	Cress et al[44]	18 Participants (44% SCI and 66% other) Setting: NR Classification: NR AIS: NR Chronicity: Chronic	D	SIP scores × WC-PFP (r=−0.38)				
WPT	Askari et al[45]	58 Participants (8.6% SCI and 91.4% other) Setting: inpatient (96%)/outpatient (4%) Classification: NR AIS: NR Chronicity: acute and subacute	V	WPT × instrumented rear wheel Spearman correlation (0.92-0.99)			A	Correlation between WPT and instrumented wheel variables: r=0.92-0.99 (+)
WST	Kirby et al[47]	24 Participants (12.5% SCI and 87.5% other) Setting: inpatient/outpatient Classification: tetraplegia/paraplegia AIS: NR Chronicity: subacute to chronic	V	Spearman WST × occupational therapist's global ratings=0.40; WST 2 × occupational therapist global rating=0.54				
	Kirby et al[48]	169 Participants (11% SCI and 89% other) Setting: inpatient/outpatient Classification: tetraplegia/paraplegia AIS: NR Chronicity: subacute to chronic	V	Spearman correlation (WST × GAWUT × admission FIM × discharge FIM) =0.39, 0.38, 0.31			V	Spearman rank correlations between WST score and the global assessments of the wheelchair users' therapists (0.394); admission FIM (0.38); discharge FIM (0.31)
	Kirby et al[46]	117 Participants (100% SCI) Setting: outpatient Classification: paraplegia/tetraplegia AIS: NR Chronicity: chronic	V	Capacity scores correlated positively with Confidence, Basic Mobility, Independence, Ability to Participate, and Satisfaction) Spearman correlation range (0.617-0.217) Performance scores correlated appropriately with Confidence, Basic Mobility, Independence, Pain, and Satisfaction Spearman correlation range (0.462-0.202)				
WST-Q	Mountain et al[50]	20 Participants (20% SCI and 80% other) Setting: inpatient Classification: NR AIS: NR Chronicity: subacute to chronic					D	Correlation between the objective WST and the WST-Q (K) (r=0.91) (+)

NOTE. Overall rating: −, insufficient; +, sufficient; ±, inconsistent; ?, indeterminate.

A, adequate; AMWC, Adapted Manual Wheelchair Circuit; AMWC+3, Adapted Manual Wheelchair Circuit+3; ATMW, Assessment Tool for Mobility in Wheelchairs; D, doubtful; ESES, Exercise Self-Efficacy Scale; FT, Functional Tasks; GSE, global score of ease; GSES, Generalized Perceived Self-efficacy Scale; I, inadequate; Imp, importance; NC, not clear; 9-WC, 9-task wheelchair Circuit; NR, not reported; OCAWUP, Obstacle Course Assessment Wheelchair User Performance; PP, paraplegia; PWFPA, Comprehensive Product-Centered Approach to Assessing Functional Performance; PWPT, Performance-Based Wheelchair Propulsion Test; Sat, satisfaction; TMWST, Timed Manual Wheelchair Slalom Test; TP, tetraplegia; V, very good; WC, Wheelchair Circuit; WT-S, Short Wheelie Test.

Table 4 Measurement properties of the included instruments: other properties

			Measurement Error		Interpretability	Responsiveness	
Instrument	Reference	Sample	Method Quality	Result (Rating)	Result (Rating)	Method Quality	Result (Rating)
PWPT	Gagnon et al[34]	14 Participants (100% SCI) Setting: inpatient Classification: tetraplegia and paraplegia (C6-L1) AIS: NR Chronicity: subacute				V	Slalom SRM=1.24/20-m propulsion max speed SRM=0.99/6-min propulsion test SRM=0.84
5AML-FIM	Middleton et al[31]	43 Participants (100% SCI) Setting: NR Classification: paraplegia/tetraplegia AIS: NR Chronicity: NC				D	Enhanced responsiveness for persons with paraplegia and poor for tetraplegia
FT	May et al[32]	20 Participants (100% SCI) Setting: NR Classification: NR AIS: NR Chronicity: NR			MDC (1: 2.27s; 2: 2.52cm; 3: 22.65s; 4: 10.64cm) SEM (1: 0.82s; 2: 0.91cm; 3: 8.17s; 4: 3.84s)		
WT-S	Fliess-Douer et al[37]	30 Participants (100% SCI) Setting: outpatient Classification: NR AIS: NR Chronicity: chronic	V	SEM: Ability (0.44)/Quality score (1.15) (+)	MDC: Ability (1.23)/Quality (3.20)		
TOWM	Fliess-Douer et al[37]	30 Participants (100% SCI) Setting: outpatient Classification: NR AIS: NR Chronicity: chronic	V	SEM (Ability 0.63)/(Quality score 1.29) (+)	MDC: Ability (1.74)/Quality (3.59)		
TMWST	Gagnon et al[40]	15 Participants (100% SCI) Setting: NR Classification: tetraplegia/paraplegia AIS: A, B, or C Chronicity: subacute to chronic	D	3.47% (+)	MDC (8.097%)		
WC	Kilkens et al[53]	27 Participants (100% SCI) Setting: NR Classification: tetraplegia/paraplegia AIS: A-D Chronicity: chronic				V	All 3 test scores had significantly improved during rehabilitation. SRM: 0.6-0.9
WhOM	Miller et al[42]	50 Participants (100% SCI) Setting: outpatient Classification: tetraplegia/paraplegia AIS: A, B, or C Chronicity: chronic	V	SEM Sat (0.43-0.58)/ Sat × Imp (5.42-5.87) (+)	MDC Sat: (1.19-1.61) Sat × Imp (15.02-16.27)		
	Alimohammad et al[43]	75 Participants (100% SCI) Setting: NR Classification: NR AIS: NR Chronicity: subacute to chronic			Ceiling effect mean Sat (22%)		

NOTE. −, insufficient; +, sufficient; ±, inconsistent; ?, indeterminate.

Abbreviations: A, adequate; AMWC, Adapted Manual Wheelchair Circuit; AMWC+3, Adapted Manual Wheelchair Circuit+3; ATMW, Assessment Tool for Mobility in Wheelchairs; D, doubtful; FT, Functional Tasks; I, inadequate; Imp, importance; MDC, minimal detectable change; NC, not clear; 9-WC, 9-Task Wheelchair Circuit; NR, not reported; OCAWUP, Obstacle Course Assessment Wheelchair User Performance; PWFPA, Comprehensive Product-Centered Approach to Assessing Functional Performance; PWPT, Performance-Based Wheelchair Propulsion Test; Sat, satisfaction; SEM, standard error of measurement; SRM, standardized response mean; TMWST, Timed Manual Wheelchair Slalom Test; V, very good; WC, Wheelchair Circuit; WT-S Short Wheelie Test.

Table 5 of the confidence level of evidence and overall rating according to COSMIN taxonomy

	Reliability			Validity							
				Content Validity			Construct Validity				
Instrument	Internal Consistency	Reliability	Measurement Error	Relevance	Comprehensiveness	Comprehensibility	Structural Validity	Hypothesis Testing	Cross-cultural/ Measurement Invariance	Criterion Validity	Responsiveness
AMWC		Low (+)						Moderate (±)		Moderate (−)	
AMWC+3										Moderate (?)	
ATMW		Low (+)									
PWFPA		Very low (−)									
5AML-FIM	Low (+)							Moderate (+)		Low (−)	Very low (−)
FT		Low (+)									
OCAWUP		Very low (+)						Low (+)			
PWPT		Low (+)						Low (+)			Low (+)
QEWS	Very low (+)	Low (+)						High (+)		Low (+)	
SEWM	Low (+)							Low (+)		Low (−)	
WT-S		Low (+)	Low (+)								
9-WC		Low (+)									
TOWM		Low (+)	Low (+)	Very Low (?)	Very low (?)	Very Low (?)		Moderate (±)			
TMWST		Very low (?)	Very low (+)								
WC								Moderate (?)			Low (+)
WhOM		Moderate (+)	Moderate (+)	Low (+)	Low (?)	Low (+)		Moderate (+)			
WC-PFP	Very low (+)	Very low (+)						Very low (−)			
WPT		Low (+)		Very Low (?)	Very Low (?)	Very Low (?)		Moderate (+)		Very Low (+)	
WST		High (+)		Moderate (?)	Moderate (?)	Moderate (?)		Moderate (±)		Low (±)	
WST-Q		Low (+)		Low (+)	Low (+)	Low (+)				Very Low (+)	
WUFA	Very low (+)	Low (+)		Very low (?)	Very low (?)	Very low (?)					

NOTE. −, insufficient; +, sufficient; ±, inconsistent; ?, indeterminate. Quality of evidence (GRADE approach): high, moderate, low, very low.

Abbreviations: AMWC, Adapted Manual Wheelchair Circuit; AMWC+3, Adapted Manual Wheelchair Circuit+3; ATMW, Assessment Tool for Mobility in Wheelchairs; FT, Functional Tasks; PWFPA, Comprehensive Product-Centered Approach to Assessing Functional Performance; 9-WC, 9-Task Wheelchair Circuit; OCAWUP, Obstacle Course Assessment Wheelchair User Performance; PWPT, Performance-Based Wheelchair Propulsion Test; TMWST, Timed Manual Wheelchair Slalom Test; WC, Wheelchair Circuit; WT-S, Short Wheelie Test.

Table 6 Recommendations for outcome measurement instrument use

	Category A		Category C	
Instrument	Sufficient Content Validity (Any Level)	At Least Low-Quality Evidence for Sufficient Internal Consistency	High-Quality Evidence for an Insufficient Measurement Property	Recommendation
AMWC	-	-	-	B
AMWC+3	-	-	-	B
ATMW	-	-	-	B
PWFPA	-	-	-	B
5AML-FIM	-	✓	-	B
FT	-	-	-	B
OCAWUP	-	-	-	B
PWPT	-	-	-	B
QEWS	-	-	-	B
SEWM	-	✓	-	B
WT-S	-	-	-	B
9-WC	-	-	-	B
TOWM	-	-	-	B
TMWST	-	-	-	B
WC	-	-	-	B
WhOM	✓	-	-	B
WC-PFP	-	-	-	B
WPT	-	-	-	B
WST	-	-	-	B
WST-Q	✓	-	-	B
WUFA	-	-	-	B

Abbreviations: AMWC, Adapted Manual Wheelchair Circuit; AMWC+3, Adapted Manual Wheelchair Circuit+3; ATMW, Assessment Tool for Mobility in Wheelchairs; FT, Functional Tasks; 9-WC, 9-task wheelchair circuit; OCAWUP, Obstacle Course Assessment Wheelchair User Performance; PWFPA, Comprehensive Product-Centered Approach to Assessing Functional Performance; PWPT, Performance-Based Wheelchair Propulsion Test; TMWST, Timed Manual Wheelchair Slalom Test; WC, Wheelchair Circuit; WT-S, Short Wheelie Test.

consistency assessed but was rated as having very low−quality evidence of indeterminate internal consistency and very low −quality evidence of inconsistency for content validity.

Reliability was the measurement property most evaluated by the studies, with 17 instruments (81%) presenting this measurement analysis. The reliability of an outcome measurement instrument is also considered important because it expresses the degree to which the measure is free from measurement error.[14,16] The WhOM and WST were rated as having moderate- and high-quality evidence of sufficient reliability, respectively. The other instruments included in this review did not show an appropriate number of published studies or data to reach moderate- or high-quality levels of evidence of essential measurement properties. According to COSMIN, these instruments were rated with the worst quality evidence and are missing several measurement properties. Furthermore, 5 instruments were rated as insufficient regarding reliability (Comprehensive Product-Centered Approach to Assessing Functional Performance: very low−quality evidence), construct validity (WC-PFP: very low−quality evidence), and criterion validity (Adapted Manual Wheelchair Circuit, 5AML-FIM, SEWM: moderate- and low-quality evidence), which lowered their recommendation level.

Of all the instruments included in the review, the WhOM and WST-Q had the essential measurement properties analyzed and obtained the highest ratings for sufficiency in several of the measurement properties. These outcome measurement instruments were rated as having low-quality evidence of sufficient content validity and low- to moderate-quality evidence of sufficient reliability and construct validity. Therefore, the WhOM and WST-Q can also be recommended as level B in clinical and research settings. However, these clinical instruments show the current best potential, among the included instruments, for the recommendation considering the quality of evidence for sufficient content validity. Despite this, further research is needed among these instruments (WhOM, WST-Q), especially regarding internal consistency, to improve their quality of evidence and to be considered recommendation level A based on the COSMIN criteria.[14]

The study sample and the setting characteristics where the measurement property analyses were performed were broad among the included studies (see tables 2-4). Specifically, the measurement properties of the WhOM and WST-Q (which are the "highly-rated" instruments in this review) were used in the outpatient setting with individuals who sustained tetraplegia or paraplegia after complete or incomplete SCI (AIS A-C) and are in the

subacute or chronic stage after injury. Additionally, the WST-Q[50] measurement properties were also analyzed in individuals with SCI classified as AIS D during inpatient rehabilitation.

The WhOM[41] and WST-Q[50] are PROMs, designed to measure manual wheelchair users' perspective of their mobility status in daily life, unlike other instruments that only measure wheelchair users' ability during a specific period of time in a controlled environment, that is, a clinical setting prepared for the test.[54] PROMs are important instruments in clinical and research contexts. They are a standardized method of quantifying patient perspective and enable appropriate patient-centered care and treatment efficacy.[54,55]

However, it the WhOM assesses self-reported items such as comfort while sitting in a wheelchair and satisfaction with the level of performance of various activities.[41–43] Although important, it may not be an optimal questionnaire for evaluating ability to maneuver obstacles, for example. On the other hand, the WST-Q was developed to assess the self-reported expected level of performance and confidence individuals have to execute specific wheelchair skills in their own setting and how often they do these skills.[46,50]

In general, the outcome measurement instruments found in this systematic review presented assessment limitations, especially in content validity, internal consistency, criterion validity, and measurement error. Very few studies with measurement properties rated as having very low− and low-quality evidence of these measurement properties were found. It is necessary to assess the responsiveness and interpretability of these outcome measurement instruments moving forward. In addition, further validation studies should be conducted with individuals with SCI who are manual wheelchair users to strengthen the quality of evidence and avoid downgrading in the GRADE evaluation.

Study strengths, limitations, and future directions

The reader should be aware that the results of this systematic review are related to the outcome measurement instrument recommendations for individuals with SCI who are manual wheelchair users. Therefore, studies with other target populations may have different results and recommendations. Additionally, only the analysis of measurement properties found in full-text articles published in indexed journals was considered in this systematic review. Although no language restrictions were applied, only studies in English were found in the consulted databases.

This systematic review has several strengths, such as a priori registered protocol, the use of a comprehensive and sensitive search strategy, and searches conducted in 5 databases (MEDLINE/PubMed, Cochrane Central Register of Controlled Trials, Web of Science, Biblioteca Regional de Medicina, and Cumulative Index to Nursing and Allied Health), the application of predefined eligibility criteria, and the use of the COSMIN, GRADE, and PRISMA guideline methodologies. Two independent reviewers carried out every review process to ensure consistency. Discrepancies were discussed and resolved by a senior researcher.

The results of this review are relevant in that they can assist clinicians and researchers in selecting the most appropriate outcome measurement instruments for assessing manual wheelchair mobility in individuals with SCI. Consequently, this will improve the ability of clinicians and researchers to appropriately measure the effectiveness and/or efficacy of rehabilitation interventions aiming at improving manual wheelchair mobility of individuals with SCI. Moreover, the review is innovative in that it was carried out using a solid methodology for conducting systematic reviews of instrument outcome measures. This review also identified the knowledge gaps that need further exploration in this population, especially regarding content validity and structural validity analysis. Further studies into the measurement properties of instruments to assess manual wheelchair mobility and their application in this population may strengthen or change existing recommendations.

Conclusions

There is a wide variety of instruments to assess mobility among manual wheelchair users living with SCI. Despite this variety, there are significant study methodological issues, and many instruments present poor-quality measurement properties that limit their recommendation for use in this population. The WhOM and the WST-Q are the current highly rated instruments. These 2 instruments are PROMs and have the current best potential to be recommended for clinical and research use to assess mobility among manual wheelchair users living with SCI.

Keywords

Rehabilitation; Spinal cord injuries; Wheelchairs; Patient outcomes assessment; Reproducibility of results

Corresponding author

Jocemar Ilha, PT, PhD, Rua Pascoal Simone, 358 Coqueiros Florianópolis, SC, Brazil CEP 88080-350. *E-mail address:* jocemar.ilha@udesc.br.

Data Availability

All data generated or analyzed during this study are included in this published article and its supplementary information files.

References

1. van der Meer P, Post MWM, van Leeuwen CMC, et al. Impact of health problems secondary to SCI one and five years after first inpatient rehabilitation. Spinal Cord 2017;55:98–104.
2. Kawano O, Maeda T, Mori E, et al. How much time is necessary to confirm the diagnosis of permanent complete cervical spinal cord injury? Spinal Cord 2020;58:284–9.
3. Post MW, van Asbeck FW, van Dijk AJ, Schrijvers AJ. Services for spinal cord injured: availability and satisfaction. Spinal Cord 1997;35:109–15.
4. Hosseini SM, Oyster ML, Kirby RL, Harrington AL, Boninger ML. Manual wheelchair skills capacity predicts quality of life and community integration in persons with spinal cord injury. Arch Phys Med Rehabil 2012;93:2237–43.
5. Kilkens OJ, Dallmeijer AJ, Nene AV, Post MW, van der Woude LH. The longitudinal relation between physical capacity and wheelchair skill performance during inpatient rehabilitation of people with spinal cord injury. Arch Phys Med Rehabil 2005;86:1575–81.
6. Morgan KA, Engsberg JR, Gray DB. Important wheelchair skills for new manual wheelchair users: health care professional and wheelchair user perspectives. Disabil Rehabil Assist Technol 2017;12:28–38.

7. Fliess-Douer O, Vanlandewijck YC, Post MWM, Van Der Woude LH, De Groot S. Wheelchair skills performance between discharge and one year after inpatient rehabilitation in hand-rim wheelchair users with spinal cord injury. J Rehabil Med 2013;45:553–9.
8. Fliess-Douer O, Vanlandewijck YC, van der Woude LH. Most essential wheeled mobility skills for daily life: an international survey among paralympic wheelchair athletes with spinal cord injury. Arch Phys Med Rehabil 2012;93:629–35.
9. van Velzen JM, van Leeuwen CMC, de Groot S, et al. Return to work five years after spinal cord injury inpatient rehabilitation: is it related to wheelchair capacity at discharge? J Rehabil Med 2012;44:73–9.
10. Fliess-Douer O, Vanlandewijck YC, Lubel Manor G, van der Woude LH. A systematic review of wheelchair skills tests for manual wheelchair users with a spinal cord injury: towards a standardized outcome measure. Clin Rehabil 2010;24:867–86.
11. Candoni G, Coronel E, Sanchez-Correa C, Tomadín R, Valdez M. [Psychometric properties of observational instruments to evaluate wheelchair skills in persons with a spinal cord injury: a systematic review] [Spanish] Rehabilitacion (Madr) 2021;55:125–37.
12. Mortenson WB, Miller WC, Auger C. Issues for the selection of wheelchair-specific activity and participation outcome measures: a review. Arch Phys Med Rehabil 2008;89:1177–86.
13. Terwee CB, Prinsen CAC, Chiarotto A, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. Qual Life Res 2018;27:1159–70.
14. Prinsen CAC, Mokkink LB, Bouter LM, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. Qual Life Res 2018;27:1147–57.
15. 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.
16. LB Mokkink, CAC Prinsen, DL Patrick, et al. COSMIN manual for systematic reviews of PROMs COSMIN methodology for systematic reviews of Patient-Reported Outcome Measures (PROMs) user manual. 2018, Amsterdam, The Netherlands.
17. Kilkens OJ, Post MWM, Dallmeijer AJ, Seelen HA, van der Woude LH. Wheelchair skills tests: a systematic review. Clin Rehabil 2003;17:418–30.
18. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71.
19. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: the PRISMA statement. J Clin Epidemiol 2009;62:1006–12.
20. Chiarotto A, Maxwell LJ, Ostelo RW, et al. Measurement properties of visual analogue scale, numeric rating scale, and Pain Severity subscale of the Brief Pain Inventory in patients with low back pain: a systematic review. J Pain 2019;20:245–63.
21. Lankhorst K, Oerbekke M, van den Berg-Emons R, Takken T, de Groot J. Instruments measuring physical activity in individuals who use a wheelchair: a systematic review of measurement properties. Arch Phys Med Rehabil 2020;101:535–52.
22. Guyatt G, Oxman AD, Akl EA, et al. GRADE guidelines: 1. Introduction—GRADE evidence profiles and summary of findings tables. J Clin Epidemiol 2011;64:383–94.
23. Schünemann H, Brożek J, Guyatt G. A Oxman. In: GRADE handbook for grading quality of evidence and strength of recommendations, The GRADE Working Group; 2013.
24. Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ 2008;336:924–6.
25. Cochrane Group. Cochrane Back Review Group. Available at: https://back.cochrane.org/. Accessed.
26. Cowan RE, Nash MS, de Groot S, van der Woude LH. Adapted manual wheelchair circuit: test-retest reliability and discriminative validity in persons with spinal cord injury. Arch Phys Med Rehabil 2011;92:1270–80.
27. Ribeiro Neto F, Costa RRG, Lopes ACG, Carregaro RL. Cross-cultural validation of a Brazilian version of the Adapted Manual Wheelchair Circuit (AMWC-Brazil). Physiother Theory Pract 2019;35:860–72.
28. Ribeiro Neto FP, Costa RRG, Carregaro RL. The addition of three new items in the Adapted Manual Wheelchair Circuit improves the discrimination between different levels of spinal cord injury. Physiother Theory Pract 2020;36:1329–39.
29. Harvey LA, Batty J, Fahey A. Reliability of a tool for assessing mobility in wheelchair-dependent paraplegics. Spinal Cord 1998;36:427–31.
30. Gil-Agudo A, Solís-Mozos M, del-Ama AJ, et al. Comparative ergonomic assessment of manual wheelchairs by paraplegic users. Disabil Rehabil Assist Technol 2013;8:305–13.
31. Middleton JW, Harvey LA, Batty J, et al. Five additional mobility and locomotor items to improve responsiveness of the FIM in wheelchair-dependent individuals with spinal cord injury. Spinal Cord 2006;44:495–504.
32. May LA, Butt C, Minor L, Kolbinson K, Tulloch K. Measurement reliability of functional tasks for persons who self-propel a manual wheelchair. Arch Phys Med Rehabil 2003;84:578–83.
33. Routhier F, Desrosiers J, Vincent C, Nadeau S. Reliability and construct validity studies of an obstacle course assessment of wheelchair user performance. Int J Rehabil Res 2005;28:49–56.
34. Gagnon DH, Roy A, Verrier MC, et al. Do performance-based wheelchair propulsion tests detect changes among manual wheelchair users with spinal cord injury during inpatient rehabilitation in Quebec? Arch Phys Med Rehabil 2016;97:1214–8.
35. Gollan EJ, Harvey LA, Simmons J, Adams R, McPhail SM. Development, reliability and validity of the Queensland Evaluation of Wheelchair Skills (QEWS). Spinal Cord 2015;53:743–9.
36. Fliess-Douer O, van der Woude LHV, Vanlandewijck YC. Development of a new scale for perceived self-efficacy in manual wheeled mobility: a pilot study. J Rehabil Med 2011;43:602–8.
37. Fliess-Douer O, van der Woude LH, Vanlandewijck YC. Reliability of the Test of Wheeled Mobility (TOWM) and the Short Wheelie Test. Arch Phys Med Rehabil 2013;94:761–70.
38. Fliess-Douer O, van der Woude LH, Vanlandewijck YC. Test of Wheeled Mobility (TOWM) and a Short Wheelie Test: a feasibility and validity study. Clin Rehabil 2013;27:527–37.
39. Kilkens OJ, Post MW, van der Woude LH, Dallmeijer AJ, van den Heuvel WJ. The wheelchair circuit: reliability of a test to assess mobility in persons with spinal cord injuries. Arch Phys Med Rehabil 2002;83:1783–8.
40. Gagnon D, Décary S, Charbonneau MF. The Timed Manual Wheelchair Slalom Test: a reliable and accurate performance-based outcome measure for individuals with spinal cord injury. Arch Phys Med Rehabil 2011;92:1339–43.
41. Mortenson WB, Miller WC, Miller-Pogar J. Measuring wheelchair intervention outcomes: development of the Wheelchair Outcome Measure. Disabil Rehabil Assist Technol 2007;2:275–85.
42. Miller WC, Garden J, Mortenson WB. Measurement properties of the wheelchair outcome measure in individuals with spinal cord injury. Spinal Cord 2011;49:995–1000.
43. Alimohammad S, Parvaneh S, Ghahari S, et al. Translation and validation of the Farsi version of the Wheelchair Outcome Measure (WhOM-Farsi) in individuals with spinal cord injury. Disabil Health J 2016;9:265–71.
44. Cress ME, Kinne S, Maher E. Physical functional performance in persons using a manual wheelchair. J Orthop Sports Phys Ther 2002;32:104–13.
45. Askari S, Kirby RL, Parker K, Thompson K, Neill JO. Wheelchair Propulsion Test: development and measurement properties of a new test for manual wheelchair users. Arch Phys Med Rehabil 2013;94:1690–8.
46. Kirby RL, Worobey LA, Cowan R, et al. Wheelchair skills capacity and performance of manual wheelchair users with spinal cord injury. Arch Phys Med Rehabil 2016;97:1761–9.

47. Kirby RL, Swuste J, Dupuis DJ, MacLeod DA, Monroe R. The Wheelchair Skills Test: a pilot study of a new outcome measure. Arch Phys Med Rehabil 2002;83:10–8.
48. RL Kirby, DJ Dupuis, AH Macphee, et al., The wheelchair skills test (version 2.4): measurement properties, *Arch Phys Med Rehabil*, **85**, 2004,794-804.
49. Lindquist NJ, Loudon PE, Magis TF, et al. Reliability of the performance and safety scores of the Wheelchair Skills Test version 4.1 for manual wheelchair users. Arch Phys Med Rehabil 2010;91:1752–7.
50. Mountain AD, Kirby RL, Smith C. The Wheelchair Skills Test, version 2.4: validity of an algorithm-based questionnaire version. Arch Phys Med Rehabil 2004;85:416–23.
51. Passuni D, Dalzotto E, Gath CF, et al. Reliability of the Spanish version of the wheelchair skills test 4.2 for manual wheelchair users with spinal cord injury. Disabil Rehabil Assist Technol 2019;14:788–91.
52. Stanley RK, Stafford DJ, Rasch E, Rodgers MM. Development of a functional assessment measure for manual wheelchair users. 2003;40:301-8.
53. Kilkens OJ, Dallmeijer AJ, de Witte LP, van der Woude LH, Post MW. The Wheelchair Circuit: construct validity and responsiveness of a test to assess manual wheelchair mobility in persons with spinal cord injury. Arch Phys Med Rehabil 2004;85:424–31.
54. Gabes M, Knüttel H, Kann G, Tischer C, Apfelbacher CJ. Measurement properties of patient-reported outcome measures (PROMs) in hyperhidrosis: a systematic review. Qual Life Res 2022;31:671–86.
55. Meadows KA. Patient-reported outcome measures: an overview. Br J Community Nurs 2011;16:146–51.