Journal of Clinical Epidemiology 149 (2022) 60−82

REVIEW ARTICLE

Multiple sclerosis disease−related knowledge measurement instruments show mixed performance: a systematic review

Marina Gonzalez-del-Rio[a,b], Carme Bertran-Noguer[c,d], Lluís Ramió-Torrentà[a,b,e], Edurne Zabaleta-del-Olmo[c,f,g,*]

[a]*Neurodegeneration and Neuroinflammation Research Group, Biomedical Research Institute (IDIBGI), Girona, Spain*
[b]*Girona Neuroimmunology and Multiple Sclerosis Unit, Neurology Department, Dr Josep Trueta Hospital and Santa Caterina Hospital, Girona-Salt, Spain*
[c]*Nursing Department, Faculty of Nursing, Universitat de Girona, Girona, Spain*
[d]*Health and Health Care Research Group, Universitat de Girona, Girona, Spain*
[e]*Medical Sciences Department, University of Girona, Girona, Spain*
[f]*Fundació Institut Universitari per a la recerca a l'Atenció Primària de Salut Jordi Gol i Gurina (IDIAPJGol), Barcelona, Spain*
[g]*Gerència Territorial de Barcelona, Institut Català de la Salut, Barcelona, Spain*

Accepted 24 May 2022; Published online 30 May 2022

Abstract

Objectives: This review aimed to summarize the evidence on the measurement properties of available disease-related knowledge measurement instruments in people with multiple sclerosis.

Study Design and Setting: We performed a literature search in the MEDLINE (PubMed), CINAHL (EBSCOhost), and PsycINFO (EBSCOhost) databases from inception to February 10, 2021. Eligible studies were reports developing a disease-related knowledge measurement instrument or assessing one or more of its measurement properties. We assessed the methodological quality of the included studies independently using the "COSMIN Risk of Bias" checklist. We graded the quality of the evidence using a GRADE approach.

Results: Twenty-four studies provided information on 14 measurement instruments. All instruments showed sufficient evidence for content validity, three for structural validity, and seven for hypothesis testing for construct validity. Cross-cultural validity and criterion validity were not assessed in any instrument. Only two instruments showed sufficient evidence for the internal consistency of their scores, and two others for their test−retest reliability. Responsiveness was assessed in one instrument, but it was rated as indeterminate.

Conclusion: Based on the available evidence, two instruments can be recommended for use, two are unrecommended, and five have the potential to be recommended for use but require further research.

Keywords: Knowledge; Multiple sclerosis; Patient-reported outcome measures; Psychometrics; Surveys and questionnaires; Systematic review

Funding: This work was supported by the Col·legi d'Infermeres/rs de Girona (Official College of Nursing of Girona) [grant number COI-GI-B01-2020]; the International Organisation of Multiple Sclerosis Nurses with the International Nightingale Award 2020; and by the Sociedad Española de Neurología (Spanish Society of Neurology) with the Neurovation Award in Multiple Sclerosis 2020. These funding sources have had no role in the design of the study, its execution, analyses, interpretation of the data, and the decision to submit results. Open Access funding provided thanks to the CRUE-CSIC agreement with Elsevier.

Conflict of interest: The authors declare that they have no conflict of interest.

Availability of data, code, and other materials: The authors confirm that the data supporting the findings of this study are available within the article and its supplementary materials.

Authors contributions: **Marina Gonzalez-del-Rio**: conceptualization; data curation; formal analysis; funding acquisition; investigation; methodology; project administration; resources; supervision; validation; visualization, and writing - original draft preparation. **Carme Bertran-Noguer**: conceptualization; formal analysis; validation; and writing - review and editing. **Lluís Ramió-Torrentà**: conceptualization; formal analysis; validation; and writing - review and editing. **Edurne Zabaleta-del-Olmo**: conceptualization; data curation; formal analysis; investigation; methodology; resources; validation; visualization; and writing - original draft preparation. All authors agreed with the results and conclusions of this manuscript. All authors had full access to all the data in the review and had final responsibility for the decision to submit for publication.

* Corresponding author. Gran Via Corts Catalanes, 587 àtic, 08007 Barcelona, Spain. Tel./fax: +34 934 824 105.

E-mail address: ezabaleta@idiapjgol.org (E. Zabaleta-del-Olmo).

https://doi.org/10.1016/j.jclinepi.2022.05.020

What is new?

Key findings

- There are evidence gaps in the available measurement instruments.
- Only 2 out of 14 identified measurement instruments are suitable.
- Five measurement instruments could potentially be recommended for use.

What this add to what is known?

- First systematic review on multiple sclerosis patient knowledge measures.

What is the implication and what should change now?

- Review findings will help make evidence-based decisions about using these instruments.

1. Introduction

Multiple sclerosis (MS) affects an estimated 2.3 million people worldwide, with a prevalence of 50−300 per 100,000 inhabitants [1]. It is considered the most common demyelinating disease and is the first cause of non−traumatic neurological disability in young adults [2]. In recent decades, the review of MS diagnostic criteria, the emergence of new therapies, and the identification of predictive biomarkers have enabled early detection and treatment of the disease [1−3]. These aspects, combined with increasingly tailored therapeutic decisions, have allowed early treatment, thus reducing relapse rates and slowing down disease progression. Moreover, an active person-centered approach [4,5] is being promoted at all stages of the disease to minimize its impact, maximize the quality of life, and adopt a wellness philosophy [1]. In this paradigm of contemplating preferences, needs and expectations of people with MS, good professional−patient communication, and shared decision-making should prevail [4−6]. To reach this optimal point, contributing to, ensuring, and improving patients' MS-related knowledge should be the first steps on this lifelong path.

In health education, knowledge is defined as the ''factual and interpretive information leading to understanding or usefulness for taking informed action'' [7]. As such, the disease-related knowledge of people with MS can influence disease self-management, coping, and adherence, which consequently affect clinical outcomes [8]. Furthermore, it is a requirement for shared decision-making, a key component of patient-centered healthcare that is critical in chronic diseases such as MS [5]. Likewise, patient knowledge is a relevant outcome to measure the effectiveness of strategies for informing, educating, and involving patients [9]. There is a need to use measurement instruments in research and practice with sufficient evidence to evaluate this outcome in a given population and context. However, several studies that evaluate the impact of information provision interventions in disease-related knowledge using measurement instruments fail to report or assess their validity [10].

Systematic reviews of outcome measurement instruments assess their quality and characteristics to determine the most suitable ones for use in clinical practice, health service planning, and research [11]. Because we found none with this approach either in the literature or in the prospective records of systematic reviews (PROSPERO database), we conducted a systematic review to summarize the evidence on the measurement properties of available disease-related knowledge measurement instruments of people with MS and identify the most suitable ones.

2. Materials and methods

This review has been conducted following the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) initiative [11−13] and reported in accordance with the Preferred Reporting Items for Systematic reviews and Meta-Analyses 2020 statement [14] and its literature search extension [15].

2.1. Eligibility criteria and information sources

Studies were eligible if they met the following criteria: (1) the instrument aimed to measure disease-related knowledge, (2) no less than 75% of the study population consisted of people with MS, and (3) the aim of the study is to develop an instrument that assessed one or more of its measurement properties or evaluated its interpretability and feasibility. Development studies of the identified instruments were also included, even if they did not involve people with MS, as such studies could provide indirect evidence on the instrument's content validity. Studies that only used the instrument as an outcome measure (e.g., clinical trials) or for validation of other instruments were excluded, as were conference abstracts. We performed a literature search without language restrictions in the MEDLINE (PubMed), CINAHL (EBSCOhost), PsycINFO (EBSCOhost), and OpenGrey and Grey Literature Report databases from their inception until February 10, 2021. We also screened the reference lists of included reports, complemented the main search with an additional one using only the instrument's name, and contacted the authors of the included studies to retrieve the maximum possible information about the instruments identified.

2.2. Search strategy and selection process

Terms in controlled language and free text were combined. Likewise, we added a highly sensitive filter

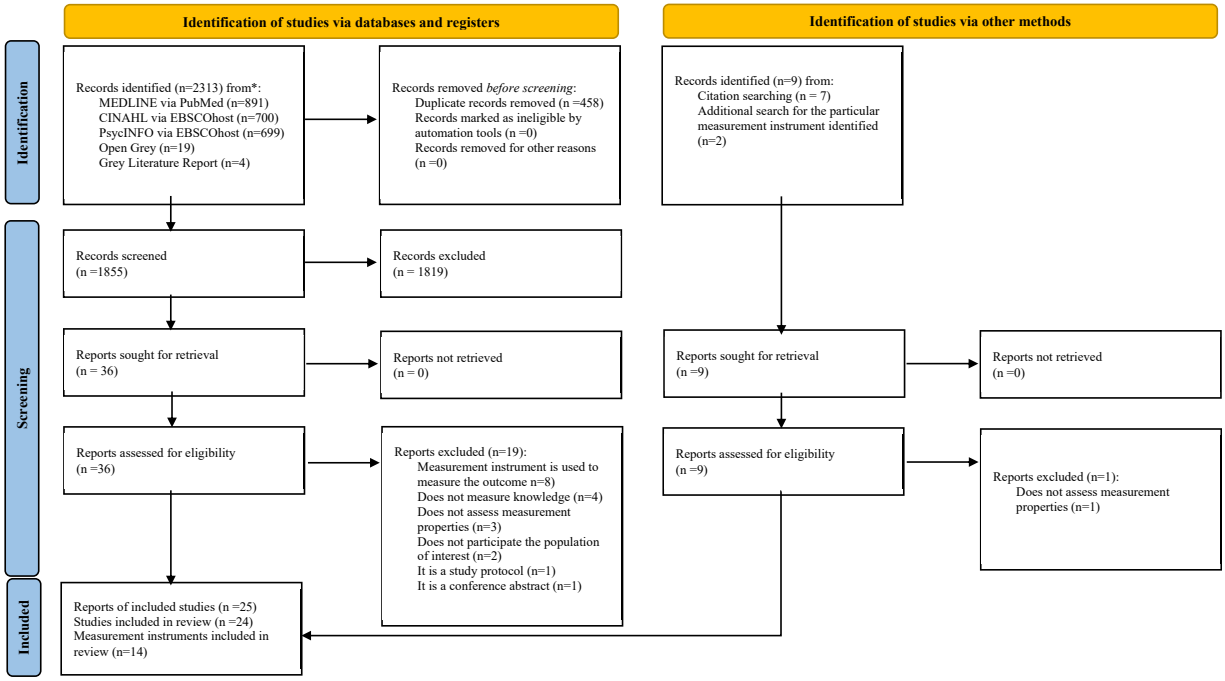

Fig. 1. PRISMA 2020 flow diagram. *Abbreviation*: PRISMA, Preferred Reporting Items for Systematic reviews and Meta-Analyses.

developed by COSMIN for the MEDLINE (PubMed) search to identify studies on the measurement properties of the instruments [16]. Reproducible searches for all databases are available at https://doi.org/10.5281/zenodo.5166552. We imported the retrieved records into the Rayyan QCRI web application program [17]. Two reviewers manually removed duplicates after using Rayyan QCRI's duplicate identification strategy. These reviewers independently screened the titles and abstracts of the records obtained, confronting them with the eligibility criteria. If a record seemed relevant to at least one of the reviewers, they independently reviewed the full-text report. Conflicts over inclusion were discussed, and a third reviewer was consulted in case of disagreement.

2.3. Data collection process and data items

We collected the information on the included studies and the identified instruments in the data extraction spreadsheets developed by COSMIN [18]. The included studies were grouped by instrument to identify the number of studies and instruments separately. A modified instrument was treated as a new instrument [11−13]. Two reviewers independently extracted the data from included studies and instruments. Disagreements were resolved by discussion between these reviewers; if no agreement could be reached, a third reviewer decided.

2.4. Study risk of bias assessment

We assessed the methodological quality of each study using the COSMIN Risk of Bias checklist. We analyzed the following measurement properties: content validity, construct validity (structural validity, hypothesis testing, and cross-cultural validity), criterion validity, reliability (internal consistency, test−retest reliability, and measurement error), and responsiveness. Concerning criterion validity, we agreed, based on the COSMIN guidelines [11,12], that no gold standard exists for identified instruments. Moreover, we did not consider *P*-values but the direction and magnitude of observed correlations in assessing hypothesis testing for construct validity and responsiveness [11−13]. The review team agreed that correlations of at least 0.50 between the instrument under study and a comparison instrument measuring the same construct, and correlations of 0.30−0.50 between instruments measuring related but different constructs, would be interpreted as adequate. Interpretability and feasibility are not considered measurement properties but are essential aspects when selecting a measurement instrument, so they were compiled in specific tables. We discussed a priori how ratings should be determined and piloted the ratings with some articles from the review to take their scope into account. Two reviewers rated all of the studies independently, which were then discussed and agreed upon by the review team.

2.5. Synthesis methods and quality of evidence

Regarding content validity, the results of each study were rated by two reviewers independently using the 10 criteria for good content validity established by the COSMIN guidelines [13]. In addition, the reviewers rated the content of the instruments themselves. Each criterion could be rated as sufficient, insufficient, or indeterminate.

Table 1. Characteristics of the identified measurement instruments

Measurement instrument (reference to the first article)	Construct(s)	Target population	Mode of administration	Recall period	(Sub)scale(s) (number of items)	Response options	Range of scores/ scoring	Original language	Available translations
Comprehension of Confidence Intervals Questionnaire (Rahn et al. 2016) [20]	Comprehension of confidence intervals	People affected by MS	Self-administration	Now	6 items	Multiple choice with one correct answer	Total score: correct answers summation Range of scores: 0–6 Higher score = higher comprehension	German	NIA
Knowledge of Multiple Sclerosis Scale (Maybury and Brewin 1984) [21]	Knowledge about MS	People affected by MS	Self-administration and interview administration	Now	14 questions	Multiple choice with multiple correct answers	Total score: summation of the assigned score (0, 1, 2) on each question Range of scores: 0–28 Higher score = higher knowledge	English (UK)	NIA
Multiple Sclerosis Health Literacy Questionnaire (Dehghani and Keshavarzi 2018) [22]	Health literacy	People affected by MS	Self-administration	Now	22 items divided into 4 subscales: (1) appraisal of health information (5 items), (2) the ability to search health information (5 items), (3) the knowledge of caring for the disease (7 items), and (4) successful practices in health conditions (5 items)	Five-point scale	Total score: summation of the assigned score (1–5) to each item Range of scores: 22–110 Higher score = higher health literacy	Persian	NIA
Multiple Sclerosis Knowledge Questionnaire (Giordano et al. 2010) [23]	Knowledge about disease	Recently diagnosed MS people	Self-administration	Now	25 items Two MSKQ versions (A and B) differ in item order, with version B intended for readministration	Multiple choice with one correct answer	Total score: correct answers summation Range of scores: 0–25 Higher score = higher knowledge	Italian	Dutch English German Turkish
Multiple Sclerosis Self-Management scale (Bishop and Frain 2007) [24]	Self-management knowledge and behaviour	People affected by MS	Self-administration	Now	38 items divided into 7 subscales: (1) treatment adherence (7 items); (2) care provider–patient relationship (5 items); (3) emotional health and social support (8 items); (4) health and symptom awareness (4 items); (5) MS knowledge and information (5 items); (6) health maintenance behaviour (5 items); and (7) communication about symptoms (4 items)	5-point Likert-type scale	Total score: summation of the assigned score (1–5) to each item Range of scores: 39–195 transformed to a scaled score (0–100) with a formula Higher score = higher self-management	English (US)	NIA

(*Continued*)

Table 1. Continued

Measurement instrument (reference to the first article)	Construct(s)	Target population	Mode of administration	Recall period	(Sub)scale(s) (number of items)	Response options	Range of scores/ scoring	Original language	Available translations
Multiple Sclerosis Self-Management scale-revised (Bishop and Frain 2011) [27]	Self-management knowledge and behaviour	People affected by MS	Self-administration	Now	24 items divided into 5 subscales: (1) healthcare provider relationship (6 items); (2) treatment adherence/ barriers (7 items); (3) social/family support (3 items); (4) MS knowledge and information (4 items); and (5) health maintenance behaviour (4 items)	5-point Likert-type scale	Total score: summation of the assigned score (1–5) to each item, three of them with a reverse score Range of scores: 24–120 transformed to a scaled score (0–100) with a formula Higher score = higher self-management	English (US)	English (Canada) Persian Polish Turkish
Multiple Sclerosis Self-Management scale-2 (Bishop et al. 2019) [34]	Self-management knowledge and behaviour	People affected by MS	Self-administration	Now	29 items divided into 7 subscales: (1) healthcare provider relationship/ communication (8 items); (2) health promotion engagement (5 items); (3) treatment adherence (3 items); (4) social/ family support (3 items); (5) MS knowledge and information (3 items), (6) health maintenance behaviour/ prevention (4 items); and (7) treatment adherence barriers (3 items)	5-point Likert-type scale	Total score: summation of the assigned score (1–5) to each item Range of scores: 29–145 transformed to a scaled score (0–100) with a formula Higher scores = higher self-management	English (US)	NIA
Patient Activation Measure (Hibbard et al. 2004) [35]	Patient activation (knowledge, skills, and confidence in self-management on health or chronic condition)	People affected by chronic conditions	Self-administration and interview administration	Now	Unidimensional scale (22 items/ statements)	Five possible responses: strongly agree, agree, disagree, strongly disagree, not applicable	Total score: summation of the assigned score (1–4) to each item, one of them with a reverse score "Not Applicable" answers: divide the score by the number of items completed and multiply by 22. Scores are not calculated for respondents who gave eight or more "Not Applicable" answers Higher scores = higher activation	English (US)	Portuguese (Brazil)

(*Continued*)

Table 1. Continued

Measurement instrument (reference to the first article)	Construct(s)	Target population	Mode of administration	Recall period	(Sub)scale(s) (number of items)	Response options	Range of scores/ scoring	Original language	Available translations
Patient Activation Measure short form (Hibbard et al. 2005) [36]	Patient activation (knowledge, skills, and confidence in self-management of one's health or chronic condition)	People affected by chronic conditions	Self-administration and interview administration	Now	Unidimensional scale (13 items/ statements)	Five possible responses: strongly agree, agree, disagree, strongly disagree, not applicable	Total score: [raw score]/[# items answered excepting nonapplicable items] × 13; can be transformed to a scale with a theoretical range 0–100 Levels of patient activation: Level 1: score under 47 Level 2: score between 47.1 and 55.2 Level 3: score between 55.2 and 67.0 Level 4: score above 67.1 Higher scores = higher activation	English (US)	Portuguese German Danish Italian Hebrew French Korean Spanish (Spain) Norwegian Swedish
Risk Knowledge questionnaire (Heesen et al. 2004) [39]	MS risk knowledge	People affected by MS	Self-administration	Now	19 items	Multiple choice with one correct answer	Total score: correct answers summation. Missing answers are considered as wrong Range of scores: 0–19 Higher score = higher knowledge	German	NIA
Risk Knowledge 1.0 questionnaire (Heesen et al. 2015) [40]	MS risk knowledge	People affected by early RRMS	Self-administration	Now	19 items divided into 5 categories: (1) general MS issues (4 items), (2) diagnosis (4 items), (3) prognosis (4 items), (4) treatment (5 items), and (5) evidence-based medicine (2 items)	Multiple choice with one correct answer	Total score: correct answers summation Range of scores: 0–19 Higher score = higher knowledge	German	NIA
Risk Knowledge 2.0 questionnaire (Heesen et al. 2017) [41]	MS risk knowledge	People affected by MS	Self-administration	Now	19 items	Multiple choice with one correct answer	Total score: correct answers summation Range of scores: 0–21 Higher score = higher knowledge	German	Dutch English (UK) Estonian Flemish French Italian Serbian Spanish (Spain) Turkish
Unnamed 1 (Abolfazli et al. 2014) [43]	Perspectives and knowledge regarding treatment	People affected by MS receiving interferon beta	Self-administration	Now	25 questions: 12 items related to knowledge and 13 to attitude	Knowledge questions: multiple choice with one correct answer Attitude questions: 5-point Likert scale	Total score: correct answers summation Range of knowledge scores: 0–12 Range of attitude scores: 13–65 Higher scores = higher knowledge and better attitudes	Persian	NIA

(*Continued*)

Table 1. Continued

Measurement instrument (reference to the first article)	Construct(s)	Target population	Mode of administration	Recall period	(Sub)scale(s) (number of items)	Response options	Range of scores/ scoring	Original language	Available translations
Unnamed 2 (Rath et al. 2017) [44]	Knowledge and understanding of risk and symptoms of PML	People affected by MS in treatment with natalizumab	Self-administration	Now	18 questions divided into 6 focus areas: (1) basic PML knowledge; (2) preinfusion questionnaire compliance; (3) wallet alert card compliance; (4) co-ownership of surveillance tests; (5) involvement desired in risk management; and (6) knowledge of other factors affecting risk	Multiple choice with a correct answer	Total score: correct answers summation Range of scores: 0–18 Higher score = higher knowledge	English (Australia)	NIA

Abbreviations: MS, multiple sclerosis; MSKQ, Multiple Sclerosis Knowledge Questionnaire; NIA, no information available; PML, progressive multifocal leukoencephalopathy; RRMS, relapsing-remitting multiple sclerosis.

Subsequently, the results of all the studies on a specific instrument and the reviewers' ratings were summarized qualitatively. The review team agreed on an overall rating of sufficient, insufficient, or inconsistent for the content validity of each instrument. For the other measurement properties, the results were rated according to the updated criteria for good measurement properties [11]. We subsequently graded the quality of their evidence using a "Grading of Recommendations Assessment, Development and Evaluation (GRADE)" approach modified by COSMIN. This approach uses four factors to determine the quality of the evidence: (1) risk of bias, (2) inconsistency of the results of the studies, (3) inaccuracy, and (4) indirect evidence. We made recommendations on the use of each identified instrument based on the available evidence and its quality grade. We classified them into three categories according to COSMIN guidelines: (A) instruments whose content validity had sufficient evidence and at least a low quality of evidence for a sufficient internal consistency of its scores, (B) instruments with high-quality evidence for an insufficient measurement property, and (C) instruments not classified either as A or B. Instruments classified as A were recommended for use while those classified as B were not. Instruments classified as C had the potential to be recommended, but further studies were needed to assess their quality.

2.6. Standard protocol approvals, registrations, and patient consents

Ethical approval and participant consent were not necessary as it is a review based solely on published studies. The review protocol was prospectively registered in the International Prospective Register of Systematic Reviews (PROSPERO CRD42019125417); no changes were made to this protocol. Details of the rationale and design of the review have been previously published [19].

3. Results

The literature search and study selection process are detailed in Fig. 1. Twenty-five reports [20–44], involving 24 studies, were included. These studies provided information about the measurement properties of 14 different instruments, whose characteristics are presented in Tables 1 and 2. The reasons for excluding reports that initially seemed to meet eligibility criteria are listed in Supplementary Text A. The results of the quality assessment of each study are shown in Tables 3 and 4. The rating of every study against the criteria for good measurement properties is described in Supplementary Table A. Finally, the summary of findings by measurement instrument is presented in Table 5.

3.1. Content validity

The development quality of nine measurement instruments was considered inadequate. This is mainly due to aspects related to the participation of the representing target population in their development and in assessing their comprehension. Furthermore, the reports of the development of some of these instruments did not describe whether a pilot test had been performed. We rated their development as doubtful for the remaining five instruments because the methodological aspects were insufficiently described. The type of instrument version assessed (final or preliminary), the interviewers' skills, the use of an interview guide, the approach used to analyze the data, and the number of

Table 2. Characteristics of the included studies

		Population					Disease/Condition characteristics		Instrument administration			
Measurement instrument	**Reference**	***N***	**Age, mean (SD, range)**	**Female**	**Education level**	**Socioeconomic characteristics**	**MS duration (yr), mean (SD)**	**Disease severity**	**Setting**	**Country**	**Language**	**Response rate**
Comprehension of Confidence Intervals Questionnaire	Rahn et al. 2016 [20]	64	IG: 47.3 yr CG: 43.8 yr	64.1%	Secondary school: 48.4% Academic degree: 51.6%	NIA	IG: 9.1 yr CG: 9.5 yr	RRMS: 65.6% SPMS: 20.3% PPMS: 3.1% CIS: 3.1%	MS day hospital Evaluative	Germany	German	55.7%
Knowledge of Multiple Sclerosis Scale	Maybury and Brewin 1984 [21]	36	42.0 yr	66.6%	NIA	Employed: 22.2%	3.7 yr	Acute phase or progressing: 19.4% Stable phase: 80.6%	Hospital and Welfare Officer Evaluative	UK	English	100.0%
MSHLQ	Dehghani and Keshavarzi 2018 [22]	210	31.9 (7.1) yr	62.8%	Under diploma: 10.9% Diploma: 16.7% Upper diploma: 72.4%	NI	7.4 (7.3) yr	RRMS: 91.9% SPMS: 0.0% PPMS: 8.1%	MS Society Evaluative	Iran	Persian	NIA
MSKQ	Giordano et al. 2010 [23]	102	35.2 (9.7) yr	67.6%	Primary: 28.4% Secondary: 54.9% College: 16.7%	Employed: 73.5% Homemaker: 11.8% Student: 10.8% Unemployed: 3.9%	NIA	RRMS: 91.2% PPMS and SPMS: 8.8%	Two MS centres Evaluative Research	Italy	Italian	97.0%
MSSM scale	Bishop and Frain 2007 [24]	266	45.7 (11.7) yr	84.6%	Below high school: 1.1% High school: 18.5% College/technical school: 38.5% College graduate: 26.4% Master's degree/ higher: 15.5%	Full time: 40.5% Part-time: 5.6% Unemployed: 5.6% Retired: 9.1% Homemakers: 7.1% Student: 3.2%	7.1 (9.3) yr	RRMS: 59.2% PPMS: 23.5% Other: 17.3%	A regional chapter of the National MS Society Evaluative	USA	English	53.0%
	Bishop et al. 2008 [25]	157	45.6 (11.3) yr	82.0%	NIA	Full time: 38.9% Part-time: 5.7% Unemployed: 55.4%	7.8 (7.7) yr	Immunotherapy: 79.0%	Two regional chapters of the National MS Society Evaluative	USA	English	39.8%
	Bishop et al. 2009 [26]	175	43.0 (9.0) yr	81.0%	High school: 18.5% College/technical school: 31.2% College graduate: 31.2% Master's degree/ higher: 11%	Full-time: 41.6% Part-time: 13.9% Unemployed: 44.5%	4.2 (3.8) yr	Immunotherapy: 80.0%	Three regional chapters of the National MS Society Evaluative	USA	English	28.0%

Table 2. Continued

		Population					Disease/Condition characteristics		Instrument administration			
Measurement instrument	**Reference**	***N***	**Age, mean (SD, range)**	**Female**	**Education level**	**Socioeconomic characteristics**	**MS duration (yr), mean (SD)**	**Disease severity**	**Setting**	**Country**	**Language**	**Response rate**
MSSM scale-revised	Bishop and Frain 2011 [27]	197	43.7 (10.4, 21–75) yr	82.7%	Under high school: 4.1% High school diploma: 17.4% College/technical school: 36.4% College graduates: 30.3% Master's degree/ higher: 11.8%	Full time: 36.9% Part-time: 11.8% Students: 3.1% Homemakers: 4.1% Retired: 7.2% Permanent disability: 30.8% Unemployed: 6.1%	3.8 (3.1) yr	RRM: 79.9% SPMS: 4.1% PPMS: 5.2% Other: 1.5% Immunotherapy: 79.0%	Three chapters of the National Multiple Sclerosis Society Evaluative	USA	English	36.0%
	Ghahari et al. 2014 [28]	31	49.4 (10.7) yr	80.6%	High school or less: 12.9% Diploma/ certificate: 19.4% College: 51.6% Master's degree/ higher: 16.1%	Full time: 25.8% Part-time: 12.9% Retired/ homemaker: 22.6% Disability pension: 45.2% Student: 6.5%	11.8 (8.0) yr	RRMS: 61.3% PPMS: 9.7% SPMS:16.1% PRMS: 3.2%	MS clinics Evaluative	Canada	English	87.1%
	Wilski et al. 2015 [29]	283	48.2 (11.8) yr	65.4%	Primary: 1.1% Vocational: 13.1% Secondary: 36.7% Higher: 49.1%	Employed: 33.2% Unemployed: 4.6% Disability pension: 47.7% Retired: 14.5%	13.5 (9.6) yr	RRMS: 33.2% PPMS: 25.1% SPMS: 22.6% PRMS: 8.8%	MS Rehabilitation Centre Evaluative	Poland	Polish	90.4%
	Wilski and Tasiemski 2016 [30]	217	47.0 (10.9) yr	66.2%	Primary/ vocational: 12.4% Secondary: 37.1% Higher: 50.5%	Monthly income of one family member 3.3%: <125€ 30.0%: 125–250€ 23.3%: 250–375€ 16.2%: 375–500€ 27.2%: >500€	12.0 (8.0) yr	RRMS: 33.8% PPMS: 24.3% SPMS: 23.8% PRMS: 6.7%	MS rehabilitation clinic Evaluative	Poland	Polish	97.0%

(*Continued*)

Table 2. Continued

Measurement instrument	Reference	Population					Disease/Condition characteristics		Instrument administration			
		N	Age, mean (SD, range)	Female	Education level	Socioeconomic characteristics	MS duration (yr), mean (SD)	Disease severity	Setting	Country	Language	Response rate
	Erbay et al. 2020 [31]	240	42.1 (10.8) yr	70.4%	Primary: 35.4% High: 30.8% University: 32.5% Postgraduate: 1.3%	NIA	Diagnosed from more than 10 yr, 88.3%	RRMS: 94.6% SPMS: 5.4%	Outpatient clinic Evaluative	Turkey	Turkish	NIA
	Saadat et al. 2020 [32]	220	35.1 (7.4) yr	69.1%	NIA	NIA	7.4 (4.4) yr	NIA	Community-based Evaluative	Iran	Persian	88.0%
	Tomczak et al. 2020 [33]	663	47.1 (11.8, 18–82) yr	66.0%	Primary/ Vocational: 19.5% Secondary: 41.6% Higher: 38.9%	Employed: 36.2% Unemployed: 5.1% Disability pension: 47.9% Retiring: 10.8%	12.3 (9.2) yr	RRMS: 37.6% PPMS: 24.0% SPMS: 20.2% PRMS: 8.7%	Rehabilitation centres and the Polish Society of MS Evaluative	Poland	Polish	NIA
MSSM-2 scale	Bishop et al. 2019 [34]	2,393	58.0 (11.3, 19–96) yr	82.9%	Under high school: 1.2% High school: 20.6% College/technical school: 28.3% College: 25.0% Master's degree/ higher: 24.3%	Full time: 18.8% Part-time: 8.4% Students: 0.4% Homemakers: 4.3% Retired: 18.6% Disability pension: 3.9%	NIA	NIA	National survey Evaluative	USA	English	34.8%
PAM	Hibbard et al. 2004 [35]	1,515	45–54 yr: 38.0% 55–64 yr: 28.0% 65–74 yr: 20.0% 75–84 yr: 13.0% 85 or older: 2.0%	63.0%	High school or less: 43.0% College or trade school: 26.0% College graduate/ higher: 31.0%	Annual household income: Less than $25,000: 32.0% $25,000 –$34,999: 12.0% $35,000 –$49,999: 17.0% $50,000 –$74.999: 17.0% $75,000 or more: 21.0%	NIA	Chronic condition: None: 21.0% Heart problem: 13.0% Arthritis: 38.0% Chronic pain: 25.0% Depression: 15.0% Diabetes: 11.0% Hypertension: 34.0% Lung disease: 12.0% Cancer: 5.0% Dyslipidemia: 30.0%	National Telephone Survey Evaluative	USA	English	48.0%

(*Continued*)

Table 2. Continued

		Population					Disease/Condition characteristics		Instrument administration			
Measurement instrument	**Reference**	***N***	**Age, mean (SD, range)**	**Female**	**Education level**	**Socioeconomic characteristics**	**MS duration (yr), mean (SD)**	**Disease severity**	**Setting**	**Country**	**Language**	**Response rate**
PAM short form	Hibbard et al. 2005 [36]	1,515	45–54 yr: 38.0% 55–64 yr: 28.0% 65–74 yr: 20.0% 75–84 yr: 13.0% 85 or older: 2.0%	63.0%	High school or less: 43.0% College or trade school: 26.0% College graduate/ higher: 31.0%	Annual household income Less than $25,000: 32.0% $25,000 –$34,999: 12.0% $35,000 –$49,999: 17.0% $50,000 –$74.999: 17.0% $75,000 or more: 21.0%	NIA	Chronic condition: None: 21.0% Heart problem: 13.0% Arthritis: 38.0% Chronic pain: 25.0% Depression: 15.0% Diabetes: 11.0% Hypertension: 34.0% Lung disease: 12.0% Cancer: 5.0% Dyslipidemia: 30.0%	National Telephone Survey Evaluative	USA	English	48.0%
	Stepleman et al. 2010 [37]/ Goodworth et al. 2016 [38]	199	46.24 (10.8) yr	82.0%	High school: 22.0% College education: 24.0% Associate's degree: 16.0% Bachelor's degree: 16.0%	Full time: 30.3% Part-time: 7.2% Unemployed: 55.9% Retired: 5.1% Disability pension: 1.5%	8.3 (6.8) yr	RRMS: 68.6% PPMS: 4.2% SPMS: 7.9% Unsure: 19.4%	MS center Evaluative	USA	English	NIA
RIKNO questionnaire	Heesen et al. 2004 [39]	169	44 (11) yr	62.7%	Higher education: 40.8%	NIA	7.7 (6.9) yr	PPMS: 50.0% RRMS: 50.0% Immunotherapy: 60.9% Early MS: 9.8%	MS outpatient clinic Evaluative	Germany	German	79.0%
RIKNO 1.0 questionnaire	Heesen et al. 2015 [40]	192	36.6 (18-70) yr	74.0%	University degree: 23.0% Secondary school: 52.0% Primary school: 25.0%	NIA	1.3 (0-2) yr	RRMS: 31.0% SPMS: 34.0% PPMS: 4.0% Immunotherapy: 45.0% Early MS: 28.0%	MS Day Hospital Evaluative	Germany	German	65.0%
RIKNO 2.0 questionnaire	Heesen et al. 2017 [41]	708	39.8 (10.2) yr	26.3%	High ≥12 yr: 52% Medium (10 –11 yr): 35%	NIA	7.1 (6.7) yr	Disability Mild: 39.0% Visible; 20.0%	MS Outpatient clinics Evaluative	Germany	German	62.1%

(*Continued*)

Table 2. Continued

Measurement instrument	Reference	Population					Disease/Condition characteristics		Instrument administration			
		N	Age, mean (SD, range)	Female	Education level	Socioeconomic characteristics	MS duration (yr), mean (SD)	Disease severity	Setting	Country	Language	Response rate
					Low (≤9 yr): 13%			Walking aids: 19.0% Wheelchair: 6.0% Early MS: 6.0% RRMS: 68.0% SPMS: 11.0% PPMS: 6.0% Immunotherapy: 65.0%				
	Giordano et al. 2018 [42]	986	38.6 (18–67) yr	77.0%	NIA	NIA	7.8 (0–37) yr	RRMS: 95.0% Immunotherapy: 77.0%	Online survey Evaluative	Germany Italy The Netherlands Serbia Spain Turkey	Dutch German Italian Serbian Spanish (Spain) Turkish	51.7%
Unnamed 1	Abolfazli et al. 2014 [43]	425	34.3 (8.4) yr	70.7%	High school: 12.2% College: 42.4% Postgraduate: 8.2%	NIA	NIA	Mean (SD) treatment with interferon beta: 37.2 (27.3) mo	Evaluative	Iran	Persian	85.0%
Unnamed 2	Rath et al. 2017 [44]	37	20–29 yr: 10.0% 30–39 yr: 26.0% 40–49 yr: 37.0% 50–59 yr: 21.0% ≥60 yr: 6.0%	67.0%	NIA	NIA	NIA	Treatment with natalizumab: 48.0%	MS-specific clinic in a major tertiary hospital Evaluative	Australia	English	77.1%

Abbreviations: SD, standard deviation; CG, control group; CIS, clinically isolated syndrome; IG, intervention group; MS, multiple sclerosis; MSHLQ, Multiple Sclerosis Health Literacy Questionnaire; MSKQ, Multiple Sclerosis Knowledge Questionnaire; MSSM, Multiple Sclerosis Self-Management; NIA, no information available; PAM, Patient Activation Measure; PPMS, primary progressive multiple sclerosis; PRMS, progressive-relapsing multiple sclerosis; RRMS, relapsing-remitting multiple sclerosis; RIKNO, Risk Knowledge; SPMS secondary progressive multiple sclerosis.

Table 3. Quality of the measurement instrument development

	Design							CI study[a]				
	General design requirements							General design requirements				
Measurement instruments	Clear construct	Clear origin of construct	Clear target population for which the instrument was developed	Clear context of use	Instrument developed in sample representing the target population	Concept elicitation[b]	Total design	CI study performed in sample representing the target population	Comprehensibility	Comprehensiveness	Total CI study	Total instrument development
Comprehension of Confidence Intervals Questionnaire [20]	V	V	V	V	I		I	V	D	D	D	I
Knowledge of Multiple Sclerosis Scale [21]	V	D	V	V	I		I				I	I
MSHLQ [22]	V	V	V	V	V	D	D	D	D	D	D	D
MSKQ [23]	V	V	V	V	V	A	A	V	D	D	D	D
MSSM scale [24]	V	V	V	V	I		I	A	D	D	D	I
MSSM scale-revised [27]	V	V	V	V	I		I				I	I
MSSM-2 scale [34]	V	V	V	V	I		I				I	I
PAM [35]	V	V	V	V	V	D	D	A	I	D	I	I
PAM short form [36]	V	V	V	V	V	D	D	A	I	D	I	I
RIKNO questionnaire [39]	V	V	V	V	D	D	D	A	D	D	D	D
RIKNO 1.0 questionnaire [40]	V	V	V	V	V	D	D	D	D	D	D	D
RIKNO 2.0 questionnaire [41]	V	V	V	V	V	D	D	V	D	D	D	D
Unnamed 1 [43]	V	V	V	V	I		I				I	I
Unnamed 2 [44]	V	V	V	V	I		I	D	D	D	D	I

Abbreviations: V, very good; A, adequate; D, doubtful; I, inadequate; CI, cognitive interview; MSHLQ, Multiple Sclerosis Health Literacy Questionnaire; MSKQ, Multiple Sclerosis Knowledge Questionnaire; MSSM, Multiple Sclerosis Self-Management; PAM, Patient Activation Measure; RIKNO, Risk Knowledge.

[a] Empty cells indicate that CI study (or part of it) was not performed.

[b] When the instrument was not developed in a sample representing the target population, the concept elicitation was not rated further.

Table 4. Quality of studies on measurement properties

	Content validity					Structural validity	Internal consistency	Cross-cultural validity	Reliability	Measurement error	Criterion validity	Construct validity		Responsiveness			
	Asking patients			Asking experts													
Instruments/ studies	Relevance	Comprehensiveness	Comprehensibility	Relevance	Comprehensiveness							Convergent validity	Known groups validity	Comparison with gold standard	Comparison with other instruments	Comparison between subgroups	Comparison before and after intervention
Comprehension of Confidence Intervals Questionnaire																	
Rahn et al. 2016 [20]							V					D					
Knowledge of Multiple Sclerosis Scale																	
Maybury and Brewin 1984 [21]												A					
MSHLQ																	
Dehghani and Keshavarzi 2018 [22]						A	V		D				D				
MSKQ																	
Giordano et al. 2010 [23]							D						D				
MSSM scale																	
Bishop and Frain 2007 [24]						A	I					V					
Bishop et al. 2008 [25]							I					V					
Bishop et al. 2009 [26]							I						D				
MSSM scale-revised																	
Bishop and Frain 2011 [27]						A	V					A					
Ghahari et al. 2014 [28]	D	D	D						A			A					
Wilski et al. 2015 [29]							I										
Wilski and Tasiemski 2016 [30]							I										

(*Continued*)

Table 4. Continued

	Content validity					Structural validity	Internal consistency	Cross-cultural validity	Reliability	Measurement error	Criterion validity	Construct validity		Responsiveness			
	Asking patients			Asking experts													
Instruments/ studies	Relevance	Comprehensiveness	Comprehensibility	Relevance	Comprehensiveness							Convergent validity	Known groups validity	Comparison with gold standard	Comparison with other instruments	Comparison between subgroups	Comparison before and after intervention
Erbay et al. 2020 [31]			I			A	V		I								
Saadat et al. 2020 [32]			I			V	V		D								
Tomczak et al. 2020 [33]			D			V	V					V					
MSSM-2 scale																	
Bishop et al. 2019 [34]						V	V					V					
PAM																	
Hibbard et al. 2004 [35]						V	V			A		I	D				
PAM short form																	
Hibbard et al. 2005 [36]						V	V						D				
Stepleman et al. 2010 [37]/ Goodworth et al. 2016 [38]						A	V					V	D				
RIKNO questionnaire																	
Heesen et al. 2004 [39]												I	D				
RIKNO 1.0 questionnaire																	
Heesen et al. 2015 [40]												D	D				I
RIKNO 2.0 questionnaire																	
Heesen et al. 2017 [41]							D					D					

(*Continued*)

Table 4. Continued

	Content validity					Structural validity	Internal consistency	Cross-cultural validity	Reliability	Measurement error	Criterion validity	Construct validity		Responsiveness			
	Asking patients			Asking experts													
Instruments/ studies	Relevance	Comprehensiveness	Comprehensibility	Relevance	Comprehensiveness							Convergent validity	Known groups validity	Comparison with gold standard	Comparison with other instruments	Comparison between subgroups	Comparison before and after intervention
Giordano et al. 2018 [42]			D									V	D				
Unnamed 1																	
Abolfazli et al. 2014 [43]							I		D								
Unnamed 2																	
Rath et al. 2017 [44]													D				

Abbreviations: V, very good; A, adequate; D, doubtful; I, inadequate; MSHLQ, Multiple Sclerosis Health Literacy Questionnaire; MSKQ, Multiple Sclerosis Knowledge Questionnaire; MSSM, Multiple Sclerosis Self-Management; PAM, Patient Activation Measure; RIKNO, Risk Knowledge.

Empty cells indicate that the measurement property assessment was not performed.

Table 5. Summary of findings for each measurement instrument according to the recommendation for use

Measurement instrument	Measurement property	Summary of results	Overall rating	Quality of evidence
Category A: Measurment instruments whose content validity had sufficient evidence, and at least a low quality of evidence for a sufficient internal consistency of its scores (recommended for use)				
PAM	Content validity	NA	Sufficient	Very low: No content validity studies, measurement instrument development study inadequate validity, and study was performed in another population of interest
	Structural validity	Rash (unidimensionality): infit and outfit ranged from $\geq$0.5 to $\leq$1.5	Sufficient	Low: There is one study of very good quality available, and study was performed in another population of interest
	Internal consistency	Cronbach's alpha 0.91; total sample size = 486	Sufficient	Low: There is one study of very good quality available, and study was performed in another population of interest
	Measurement error	MIC not defined	Indeterminate	
	Hypothesis testing	5 out of 6 hypotheses confirmed	Sufficient	Very low: There is one study of doubtful quality available, and study was performed in another population of interest
PAM short form	Content validity	NA	Sufficient	Very low: No content validity studies, measurement instrument development study inadequate validity, and only part of the study population consisted of patients with the disease of interest
	Structural validity	Rash (unidimensionality): infit and outfit ranged from $\geq$0.5 to $\leq$1.5	Sufficient	Low: There is one study of very good quality available, inconsistency was found, and only part of the study population consisted of patients with the disease of interest
	Internal consistency	Cronbach's alpha 0.88; total sample size = 1,714	Sufficient	Moderate: There are two studies of very good quality available and only part of the study population consisted of patients with the disease of interest
	Hypothesis testing	11 out of 11 hypotheses confirmed	Sufficient	Moderate: There is one study of very good quality available and only part of the study population consisted of patients with the disease of interest
Category B: Measurement instruments with high-quality evidence for an insufficient measurement property (unrecommended for use)				
MSSM scale-revised	Content validity	NA	Sufficient	Moderate: At least one content validity study of doubtful quality

(*Continued*)

Table 5. Continued

Measurement instrument	Measurement property	Summary of results	Overall rating	Quality of evidence
	Structural validity	Results of CFAs are inconsistent	Inconsistent	
	Internal consistency	Cronbach's alpha 0.59–0.91; total sample size = 1,616	Indeterminate	
	Reliability	ICC 0.64–0.88; total sample size = 31	Sufficient	Very low: There is one study of adequate quality available, and the total sample included in the study is below 50
	Hypothesis testing	3 out of 7 hypotheses confirmed	Insufficient	High: There is one study of very good quality available
MSSM-2 scale	Content validity	NA	Sufficient	Very low: No content validity studies and measurement instrument development study inadequate validity
	Structural validity	CFA (7 factors): CFI = 0.91 and RMSEA = 0.05	Sufficient	High: There is one study of very good quality available
	Internal consistency	Cronbach's alpha 0.54–0.89; total sample size = 1,197	Insufficient	High: There is one study of very good quality available
	Hypothesis testing	1 out of 5 hypotheses confirmed	Insufficient	High: There is one study of very good quality available
Category C: Measurement instruments categorized not in A or B (recommended for use until further evidence is provided)				
Comprehension of Confidence Intervals Questionnaire	Content validity	NA	Sufficient	Very low: No content validity studies and measurement instrument development study inadequate validity
	Internal consistency	Cronbach's alpha 0.57–0.21; total sample size = 64	Indeterminate	
	Hypothesis testing	0 out of 1 hypothesis confirmed	Insufficient	Very low: There is only one study of doubtful quality available, and the total sample included in the study is below 100
Knowledge of Multiple Sclerosis Scale	Content validity	NA	Sufficient	Very low: No content validity studies and measurement instrument development study inadequate validity
	Hypothesis testing	There is no information about results	Indeterminate	
MSHLQ	Content validity	NA	Sufficient	Low: No content validity studies and measurement instrument development study doubtful validity
	Structural validity	EFA: 58% of explained variance	Indeterminate	
	Internal consistency	Cronbach's alpha 0.84–0.97; total sample size = 210	Indeterminate	

(*Continued*)

Table 5. Continued

Measurement instrument	Measurement property	Summary of results	Overall rating	Quality of evidence
	Reliability	ICC 0.88–0.96; total sample size = 20	Sufficient	Very low: There is only one study of doubtful quality available, and the total sample included in the study is below 50
	Hypothesis testing	1 out of 1 hypothesis confirmed	Sufficient	Low: There is only one study of doubtful quality available
MSKQ	Content validity	NA	Sufficient	Low: No content validity studies and measurement instrument development study doubtful validity
	Internal consistency	KR20: 0.76; total sample size = 102	Indeterminate	
	Hypothesis testing	2 out of 2 hypotheses confirmed	Sufficient	Low: There is only one study of doubtful quality available
MSSM scale	Content validity	NA	Sufficient	Very low: No content validity studies and measurement instrument development study inadequate validity
	Structural validity	EFA: 50% of explained variance	Indeterminate	
	Internal consistency	Cronbach's alpha 0.84 –0.87; total sample size = 554	Indeterminate	
	Hypothesis testing	6 out of 8 hypotheses confirmed	Sufficient	High: There are two studies of very good quality available
RIKNO questionnaire	Content validity	NA	Sufficient	Low: No content validity studies and measurement instrument development study doubtful validity
	Hypothesis testing	3 out 4 hypotheses confirmed	Sufficient	Low: There is only one study of doubtful quality available
RIKNO 1.0 questionnaire	Content validity	NA	Sufficient	Low: No content validity studies and measurement instrument development study doubtful validity
	Hypothesis testing	0 out 10 hypotheses confirmed	Insufficient	Low: There is only one study of doubtful quality available
	Responsiveness	No hypothesis defined	Indeterminate	
RIKNO 2.0 questionnaire	Content validity	NA	Sufficient	Low (study on the relevance and comprehensiveness): No content validity studies and measurement instrument development study doubtful validity Moderate (study on the comprehensibility): At least one content validity study of doubtful quality

(*Continued*)

Table 5. Continued

Measurement instrument	Measurement property	Summary of results	Overall rating	Quality of evidence
	Internal consistency	Cronbach's alpha 0.73; total sample size = 708	Indeterminate	
	Hypothesis testing	4 out 4 hypotheses confirmed	Sufficient	High: There is one study of very good quality available
Unnamed 1	Content validity	NA	Sufficient	Very low: No content validity studies and measurement instrument development study inadequate validity
	Internal consistency	Not all information for "+" reported; total sample size = 20	Indeterminate	
	Reliability	Not all information for "+" reported; total sample size = 20	Indeterminate	
Unnamed 2	Content validity	NA	Sufficient	Very low: No content validity studies and measurement instrument development study inadequate validity
	Hypothesis testing	5 out 7 hypotheses confirmed	Insufficient	Very low: There is only one study of doubtful quality available, and the total sample size included in the study is below 50

Abbreviations: CFA, confirmatory factor analysis; CFI, comparative fit index; EFA, exploratory factor analysis; ICC, intraclass correlation coefficient; KR, Kuder–Richardson; MIC, minimal important change; MSHLQ, Multiple Sclerosis Health Literacy Questionnaire; MSKQ, Multiple Sclerosis Knowledge Questionnaire; MSSM, Multiple Sclerosis Self-Management; NA, not applicable; PAM, Patient Activation Measure; RIKNO, Risk Knowledge; RMSEA, root mean square error of approximation.

researchers involved in the analysis were some of the unclear aspects.

We found four studies that assessed content validity aspects for the Multiple Sclerosis Self-Management (MSSM) scale-revised and one that assessed them for the RIKNO (Risk Knowledge) 2.0 questionnaire (Table 4). Ghahari et al. [28] assessed the relevance, comprehensiveness, and comprehensibility of the items of the MSSM scale-revised. However, the method used was not clearly described. The comprehensibility of Persian [32], Polish [33], and Turkish [31] versions of this instrument was also assessed, but either the target population did not participate in the assessment [31,32], or the methodology was not clearly described [33]. Regarding the RIKNO 2.0 questionnaire, the content validity study only assessed the comprehensibility of its translated versions.

Thus, all instruments showed sufficient evidence for content validity. However, due to the low methodological quality of development studies, which were predominantly inadequate or doubtful, and the scarcity of content validity studies, the reviewers' ratings mainly counted for the evidence synthesis, leading to very low or low quality evidence of sufficient content validity for most instruments (Table 5).

3.2. Construct validity

We did not identify any studies that assessed the structural validity of eight instruments. The structural validity of two instruments [Multiple Sclerosis Health Literacy Questionnaire (MSHLQ) and MSSM scale] was rated as indeterminate because the identified studies did not provide sufficient information. Findings on the structural validity of the MSSM scale-revised were inconsistent. The confirmatory factor analysis of a five-factor structure showed a good model fit in one study [32] and a poor model fit in another [33] (Supplementary Table A). Only three instruments showed sufficient evidence for structural validity: the MSSM-2 scale, the Patient Activation Measure (PAM), and the PAM short form (Table 5). Hypothesis testing for construct validity was assessed in 13 instruments. The results of the convergent validity and discriminative validity tests are described in Supplementary Table A. Seven instruments showed sufficient evidence, five showed insufficient evidence for this measurement property, and one could not be rated due to a lack of information. We did not find any study that assessed cross-cultural validity aspects.

3.3. Criterion validity

None of the included studies reported a comparison of a shortened instrument with its original long version.

3.4. Reliability

The internal consistency of the scores was assessed in nine instruments (Supplementary Table A and Table 5). However, we could only determine the results of three instruments as the rest presented insufficient evidence on their structural validity. The PAM and the PAM short form showed sufficient evidence for this measurement property with low and moderate quality, respectively. In contrast, the MSSM-2 scale showed insufficient evidence with a high degree of quality. Test−retest reliability of scores was assessed in three instruments. The MSHLQ and the MSSM scale-revised showed sufficient evidence for this measurement property, but both with very low quality (Table 5). Regarding the instrument developed by Abolfazli et al. [43], the available evidence for this property could not be rated because they did not provide information on the observed intraclass correlation coefficients. Measurement error was only assessed in the PAM short form, but we could not interpret this given the unavailability of information on the minimal important change.

3.5. Responsiveness

The responsiveness of scores was assessed only in the RIKNO 1.0 questionnaire (Supplementary Table A and Table 5). However, the statistical significance of the change was assessed rather than testing hypotheses about expected differences in changes between the groups.

3.6. Categorization of measurement instruments

Based on the findings, we classified two instruments as A, the PAM and the PAM short form, and these can be recommended for use. The MSSM scale-revised and the MSSM-2 scale were classified as B, so their use cannot be recommended. The remaining 10 instruments were categorized as C; the quality of the content validity evidence for 5 of them was higher than the others [the MSHLQ, the Multiple Sclerosis Knowledge Questionnaire (MSKQ), and the three RIKNO questionnaire versions] (Table 5), and they could be provisionally recommended for use until further evidence is provided. The RIKNO questionnaire versions have the fewest number of items; however, a certain level of numeracy may be required from participants to answer it (Supplementary Table B). The length of the MSHLQ and the MSKQ is similar, although the completion time was shorter in the case of the MSHLQ; furthermore, the instruments' scores were positively associated with the educational level of the participants.

4. Discussion

This review was designed to determine the most suitable measurement instruments of disease-related knowledge of people with MS. Its findings show that only two instruments can be recommended for use, and five could be provisionally recommended until further evidence is provided.

Comprehensive database searches and the use of a rigorous and innovative methodology are key strengths of this review. However, subjectivity may have affected review processes. Studies were reviewed independently to mitigate this potential limitation and ratings were agreed upon by consensus among the review team to reduce interpretation variability. In addition, psychometric reviews are complex as they involve multiple reviews, one for each measurement property. Consequently, the review team included reviewers with knowledge of the construct of interest and experience with the target population and with the field of psychometrics and qualitative research. Finally, public and patient involvement in research is an increasingly important issue [45]. Therefore, patient involvement in future reviews would probably have to be considered, particularly in assessing the content validity of the instruments.

We found that the PAM and the PAM short form were the most suitable instruments. Both measure "activation," understanding this construct as knowledge, skills, and confidence in self-management of one's health or chronic condition [35,36]. Therefore, they do not strictly measure knowledge but consider it a subconstruct of the "activation." In addition, both are generic instruments, not specifically targeted at people with MS, although they have been used in studies in the area of MS [46,47]. However, researchers and clinicians should consider other instruments to perform more specific measures of MS-related knowledge.

Among these specific instruments, we provisionally recommended the use of five until more evidence is available: the MSHLQ, the MSKQ, and the three RIKNO questionnaire versions. The quality of the evidence for their content validity was rated as low. We did not identify any content validity studies of these five instruments other than one conducted by Giordano et al. [42] to assess the comprehensibility of the RIKNO 2.0 questionnaire translation. Furthermore, we only identified a single study that assessed four of these instruments [22,23,39,40] and two studies that assessed the RIKNO 2.0 version [41,42]. Given that measurement instruments are used with diverse groups of people and in different circumstances, further evidence is needed to assess whether they are valid and reliable for such use [11,13].

According to the COSMIN guidelines, a measurement instrument can be recommended for use if, in addition to sufficient evidence of content validity, it shows sufficient internal consistency of at least low quality evidence [11−13]. Sufficient evidence of the structural validity of the measurement instrument is needed to be able to interpret

the internal consistency coefficients [11]. Therefore, the evidence on the internal consistency of seven instruments was rated as indeterminate due to the lack of evidence of this measurement property. Furthermore, analysis of the internal structure is only relevant when the instrument is based on a reflective model that assumes all items of a scale or subscale are manifestations of an underlying construct [48]. None of the included studies described the type of model on which they were based. Consequently, according to COSMIN's recommendations, we considered all the identified instruments to be based on a reflective model and interpreted the analyses of their internal structure [12]. In future analyses of these instruments and the development of new ones, it would be desirable to report whether instruments are based on reflective or formative models to justify the relevance of structural validity analysis.

The most studied measurement property, and for which seven of the instruments showed sufficient evidence, was hypothesis testing for construct validity. On the other hand, the test−retest reliability was one of the least studied: only three instruments performed such an assessment, and only two presented sufficient evidence, which was of very low quality. Concerning cross-cultural validity, although many original versions have been translated into other languages or adapted to other cultures, we have not identified any studies that have assessed this. Such studies are necessary to assess whether measures from one population of a given culture are equivalent to those from another population with different cultural characteristics [12].

Based on the available evidence, only 2 out of 14 disease-related knowledge measurement instruments are suitable. However, these two instruments assess a broader construct than knowledge and are not explicitly aimed at people with MS. Five instruments could potentially be recommended for use among the identified instruments that strictly measure MS-related knowledge. Nevertheless, further research is required to examine their suitability more closely. This review identifies evidence gaps in the available measurement instruments and thus provides a helpful framework for both new assessments of these instruments and the development of new ones. Review findings will also help researchers and clinicians make evidence-based decisions about the use of these measurement instruments.

Acknowledgment

The authors appreciate the review of the English text by Mr Andy Hughes.

Appendix B

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jclinepi.2022.05.020.

References

[1] Thompson AJ, Baranzini SE, Geurts J, Hemmer B, Ciccarelli O. Multiple sclerosis. Lancet 2018;391(10130):1622−36.
[2] Filippi M, Bar-Or A, Piehl F, Preziosa P, Solari A, Vukusic S, et al. Multiple sclerosis. Nat Rev Dis Primer 2018;4(1):43.
[3] Wingerchuk DM, Weinshenker BG. Disease modifying therapies for relapsing multiple sclerosis. BMJ 2016;354:i3518.
[4] Bokhour BG, Fix GM, Mueller NM, Barker AM, Lavela SL, Hill JN, et al. How can healthcare organizations implement patient-centered care? Examining a large-scale cultural transformation. BMC Health Serv Res 2018;18:168.
[5] Heesen C, Köpke S, Solari A, Geiger F, Kasper J. Patient autonomy in multiple sclerosis — possible goals and assessment strategies. J Neurol Sci 2013;331(1−2):2−9.
[6] Colligan E, Metzler A, Tiryaki E. Shared decision-making in multiple sclerosis. Mult Scler J 2017;23(2):185−90.
[7] Communication theory and health behavior change. In: Glanz K, Rimer BK, Viswanath K, editors. Health behavior and health education: theory, research, and practice. 4th ed. San Francisco, CA: Jossey-Bass; 2008.
[8] Vermersch P, Shanahan J, Langdon D, Yeandle D, Alexandri N, Schippling S. Knowledge is power, but is ignorance bliss? Optimising conversations about disease progression in multiple sclerosis. Neurol Ther 2020;9(1):1−10.
[9] Coulter A, Ellins J. Effectiveness of strategies for informing, educating, and involving patients. BMJ 2007;335:24−7.
[10] Köpke S, Solari A, Rahn A, Khan F, Heesen C, Giordano A. Information provision for people with multiple sclerosis. Cochrane multiple sclerosis and rare diseases of the CNS group. Cochrane Database Syst Rev 2018CD008757.
[11] Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. Qual Life Res 2018;27:1147−57.
[12] Mokkink LB, de Vet HCW, Prinsen CAC, Patrick DL, Alonso J, Bouter LM, et al. COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. Qual Life Res 2018;27:1171−9.
[13] Terwee CB, Prinsen CAC, Chiarotto A, Westerman MJ, Patrick DL, Alonso J, 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] Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71.
[15] Rethlefsen ML, Kirtley S, Waffenschmidt S, Waffenschmidt S, Ayala AP, Moher D, et al. PRISMA-S: an extension to the PRISMA statement for reporting literature searches in systematic reviews. Syst Rev 2021;10(1):39.
[16] Terwee CB, Jansma EP, Riphagen II, de Vet HCW. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. Qual Life Res 2009;18:1115−23.
[17] Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan—a web and mobile app for systematic reviews. Syst Rev 2016;5(1):210.
[18] COnsensus-based Standards for the selection of health Measurement INstruments. COSMIN tools. COSMIN. Available at https://www.cosmin.nl/tools/guideline-conducting-systematic-review-outcome-measures/?portfolioCats=19. Accessed February 28, 2022.
[19] Gonzalez-del-Rio M, Bertran-Noguer C, Ramió-Torrentà L, Zabaleta-del-Olmo E. Disease-related knowledge measurement instruments of people affected by multiple sclerosis: protocol for a systematic psychometric review. BMJ Open 2021;11(10):e049580.
[20] Rahn AC, Backhus I, Fuest F, Riemann-Lorenz K, Köpke S, van de Roemer A, et al. Comprehension of confidence intervals - development and piloting of patient information materials for people with

multiple sclerosis: qualitative study and pilot randomised controlled trial. BMC Med Inform Decis Mak 2016;16(1):122.
[21] Maybury CP, Brewin CR. Social relationships, knowledge and adjustment to multiple sclerosis. J Neurol Neurosurg Psychiatry 1984; 47(4):372−6.
[22] Dehghani A, Keshavarzi A. Development and validation of a multidimensional health literacy questionnaire for multiple sclerosis patients. Mult Scler Relat Disord 2018;25:156−62.
[23] Giordano A, Messmer Uccelli M, Pucci E, Martinelli V, Borreani C, Lugaresi A, et al. The Multiple Sclerosis Knowledge Questionnaire: a self-administered instrument for recently diagnosed patients. Mult Scler 2010;16(1):100−11.
[24] Bishop M, Frain M. Development and initial analysis of multiple sclerosis self-management scale. Int J MS Care 2007; 9(2):35−42.
[25] Bishop M, Frain MP, Tschopp MK. Self-management, perceived control, and subjective quality of life in multiple sclerosis: an exploratory study. Rehabilitation Counseling Bulletin 2008;52(1):45−56.
[26] Bishop M, Frain MP, Rumrill PD, Rymond C. The relationship of self-management and disease modifying therapy use to employment status among adults with multiple sclerosis. J Vocat Rehabil 2009; 31(2):119−27.
[27] Bishop M, Frain MP. The multiple sclerosis self-management scale: revision and psychometric analysis. Rehabil Psychol 2011;56(2): 150−9.
[28] Ghahari S, Khoshbin LS, Forwell SJ. The multiple sclerosis self-management scale. Int J MS Care 2014;16(2):61−7.
[29] Wilski M, Tasiemski T, Kocur P. Demographic, socioeconomic and clinical correlates of self-management in multiple sclerosis. Disabil Rehabil 2015;37(21):1970−5.
[30] Wilski M, Tasiemski T. Illness perception, treatment beliefs, self-esteem, and self-efficacy as correlates of self-management in multiple sclerosis. Acta Neurol Scand 2016;133(5):338−45.
[31] Erbay Ö, Usta Yeşilbalkan Ö, Yüceyar N, Baklan M, Karadakovan A, Tekindal MA. Validity and reliability study of the Turkish version of multiple sclerosis self-management scale. J Neurosci Nurs 2020; 52(3):122−7.
[32] Saadat S, Kajbaf MB, Kalantari M, Hosseininezhad M. The multiple sclerosis self-management scale−revised (MSSM-R). Int J MS Care 2020;22(1):37−42.
[33] Tomczak M, Kleka P, Wilski M. Psychometric properties of the polish version of the multiple sclerosis self-management scale − revised. Disabil Rehabil 2020;44(10):2113−22.
[34] Bishop ML, Frain MP, Li J, Chiu CY, McDaniels B, Kim BJ. The multiple sclerosis self-management scale-2: evaluation of an updated scale. J Appl Rehabil Couns 2019;50(3):210−26.
[35] Hibbard JH, Stockard J, Mahoney ER, Tusler M. Development of the patient activation measure (PAM): conceptualizing and measuring activation in patients and consumers: development of the patient activation measure (PAM). Health Serv Res 2004;39:1005−26.
[36] Hibbard JH, Mahoney ER, Stockard J, Tusler M. Development and testing of a short form of the patient activation measure. Health Serv Res 2005;40:1918−30.
[37] Stepleman L, Rutter MC, Hibbard J, Johns L, Wright D, Hughes M. Validation of the patient activation measure in a multiple sclerosis clinic sample and implications for care. Disabil Rehabil 2010; 32(19):1558−67.
[38] Goodworth MCR, Stepleman L, Hibbard J, Johns L, Wright D, Hughes MD, et al. Variables associated with patient activation in persons with multiple sclerosis. J Health Psychol 2016;21(1): 82−92.
[39] Heesen C, Kasper J, Segal J, Köpke S, Mühlhauser I. Decisional role preferences, risk knowledge and information interests in patients with multiple sclerosis. Mult Scler J 2004;10(6):643−50.
[40] Heesen C, Kasper J, Fischer K, Köpke S, Rahn A, Backhus I, et al. Risk knowledge in relapsing multiple sclerosis (RIKNO 1.0) - development of an outcome instrument for educational interventions. In: Meuth SG, editor. PLoS One 2015;10:e0138364.
[41] Heesen C, Pöttgen J, Rahn AC, Liethmann K, Kasper J, Vahter L, et al. What should a person with relapsing-remitting multiple sclerosis know? − focus group and survey data of a risk knowledge questionnaire (RIKNO 2.0). Mult Scler Relat Disord 2017;18:186−95.
[42] Giordano A, Liethmann K, Köpke S, Poettgen J, Rahn AC, Drulovic J, et al. Risk knowledge of people with relapsing-remitting multiple sclerosis − results of an international survey. In: Aktas O, editor. PLoS One 2018;13:e0208004.
[43] Abolfazli R, Elyasi A, Javadi MR, Gholami K, Torkamandi H, Amir-Shahkarami M, et al. Knowledge and attitude assessment of Iranian multiple sclerosis patients receiving interferon beta. Iran J Neurol 2014;13(3):160−7.
[44] Rath L, Vijiaratnam N, Skibina O. Assessing understanding of individual risk and symptoms of progressive multifocal leukoencephalopathy in patients prescribed natalizumab for multiple sclerosis: patient understanding: natalizumab-PML. Intern Med J 2017;47(2):194−9.
[45] Domecq JP, Prutsky G, Elraiyah T, Wang Z, Nabhan M, Shippee N, et al. Patient engagement in research: a systematic review. BMC Health Serv Res 2014;14:89.
[46] Wilkie DD, Solari A, Nicholas RSJ. The impact of the face-to-face consultation on decisional conflict in complex decision-making in multiple sclerosis: a pilot study. Mult Scler J Exp Transl Clin 2020;6(4). 205521732095980.
[47] Rahn AC, Wenzel L, Icks A, Stahmann A, Scheiderbauer J, Grentzenberg K, et al. Development and evaluation of an interactive web-based decision-making programme on relapse management for people with multiple sclerosis (POWER@MS2) —study protocol for a randomised controlled trial. Trials 2021;22(1):139.
[48] Avila ML, Stinson J, Kiss A, Brandão LR, Uleryk E, Feldman BM. A critical review of scoring options for clinical measurement tools. BMC Res Notes 2015;8(1):612.