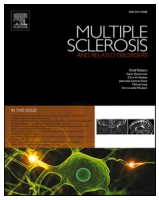
Multiple Sclerosis and Related Disorders 69 (2023) 104461

Contents lists available at ScienceDirect

Multiple Sclerosis and Related Disorders

journal homepage: www.elsevier.com/locate/msard



Original article

Development of an international, multidisciplinary, patient-centered Standard Outcome Set for Multiple Sclerosis: The S.O.S.MS project

K. Daniels [a,b,*], S.T.F.M. Frequin [c], E.M.W. van de Garde [d,e], D.H. Biesma [f,g], P.J. van der Wees [b], P.B. van der Nat [a,b], (S.O.S.MS Consensus Group)

[a] Department of Value-Based Healthcare, St. Antonius Hospital, Utrecht/Nieuwegein, the Netherlands
[b] Radboud university medical center, Radboud Institute for Health Sciences, Scientific Center for Quality of Healthcare (IQ healthcare), the Netherlands
[c] Department of Neurology, St. Antonius Hospital, Utrecht/Nieuwegein, the Netherlands
[d] Department of Clinical Pharmacy, St. Antonius Hospital, Utrecht/Nieuwegein, the Netherlands
[e] Division of Pharmacoepidemiology and Clinical Pharmacology, Faculty of Science, Utrecht University, Utrecht, the Netherlands
[f] Department of Internal Medicine, Leiden University Medical Center, Leiden, the Netherlands
[g] Department of Internal Medicine, University Medical Centre Utrecht, Utrecht, the Netherlands

ARTICLE INFO

Keywords:
Outcome set
Core outcomes
Value-based healthcare
Benchmark
Outcome indicator
Patient-reported outcomes
PROM
Multiple sclerosis

ABSTRACT

Background: Currently, outcomes of Multiple Sclerosis (MS) are not standardized and it is unclear which outcomes matter most to people living with MS. A consensus between patients and healthcare professionals on which outcomes to measure and how, would facilitate a move towards value-based MS care.
Objective: to develop an internationally accepted, patient-relevant Standard Outcome Set for MS (S.O.S.MS).
Methods: A mixed-method design was used, including a systematic literature review, four patient focus groups (n=30) and a RAND-modified Delphi process with seventeen MS experts of five disciplines from seven countries (the Netherlands, United States of America, Portugal, Ireland, India, New Zealand, Switzerland and Turkey).
Results: A standard outcome set for MS was defined, consisting of fourteen outcomes divided in four domains: disease activity (n=3), symptoms (n=4), functional status (n=6), and quality of life (n=1). For each outcome, an outcome measure was selected and the measurement protocol was defined. In addition, seven case-mix variables were selected.
Conclusion: This standard outcome set provides a guideline for measuring outcomes of MS in clinical practice and research. Using this set to monitor and (inter)nationally benchmark real-world outcomes of MS can support improvement of patient value and ultimately guide the transition towards value-based MS care.

1. Introduction

Currently, there is no consensus on which treatment outcomes need to be measured in the care for people with Multiple Sclerosis (MS). Furthermore, it is relatively unknown which outcomes matter most to people living with MS. Insight in real-world outcomes of MS is still limited and practice variation can be found between countries and centers (Marziniak et al., 2016). The lack of insight in real-world outcomes is especially important in light of the many new disease modifying treatments (DMTs) that have been discovered and made available to patients with MS over the last decade.

Monitoring and (inter)nationally comparing real-world outcomes of care provides the opportunity to improve patients' value, defined as outcomes achieved per monetary unit spent (Porter, 2010). However, in order to move towards value-based MS care, there is a need for global standardization in the measurement of MS outcomes. A standard outcome set can support outcome improvement for instance by learning from practice variations, comparing outcomes between healthcare providers, and informing patients of treatment outcomes.

Standard outcome sets have been developed for other chronic diseases (Verberne et al., 2019, Kampstra et al., 2019) and neurological conditions (de Roos et al., 2017, Salinas et al., 2016). For MS, previous studies attempted to standardize outcomes (Paul et al., 2014, Cohen et al., 2015, Potter et al., 2014, Swart et al., 2019, van Munster and Uitdehaag, 2017), but these studies targeted specific treatment groups such as exercise programs (Paul et al., 2014) and rehabilitation

* Corresponding author at: St. Antonius hospital, Department of value-based healthcare, P.O. Box 2500, 3430 EM Nieuwegein, the Netherlands.
E-mail address: k.daniels@antoniusziekenhuis.nl (K. Daniels).

https://doi.org/10.1016/j.msard.2022.104461
Received 15 July 2021; Received in revised form 15 November 2022; Accepted 7 December 2022
Available online 16 December 2022

programs (Cohen et al., 2015, Potter et al., 2014), or had a specific aim such as value-based contracting (Swart et al., 2019) or clinical trials (van Munster and Uitdehaag, 2017). None of these initiatives produced a holistic standard set with multiple domains, that was focused on what matters most to patients and that is globally applicable for all MS-care providers. Therefore, the aim of this study was to develop an internationally accepted, multidisciplinary, patient-relevant standard outcome set for MS.

2. Method

This study was registered with the COMET Initiative (www.comet-initiative.org) and consisted of three parts: a) literature review, b) patient focus groups and c) Rand-modified Delphi process with a multidisciplinary expert panel (Fig. 1). The results of the literature review and the patient focus groups served as input for the Rand-modified Delphi process.

The framework of Valderas and Alonso (2008) was used to classify the outcomes and outcome measures into domains. According to Valderas and Alonso, outcomes can be grouped into multiple constructs: disease activity (biological), symptoms, functional status, health perceptions, health-related quality of life and other health related constructs such as 'satisfaction with care'.

a) Literature review

The literature review was performed and designed by the research team (KD and STFMF) in consultation with a clinical librarian (Appendix A). The search was performed using PubMed, PsychINFO and Embase. Since the literature review served exclusively as input for the consensus meetings, a saturation method was used. First, 200 references were randomly selected from the search results and screened on eligibility (title/abstract) by the reviewer (KD). Second, data (i.e. outcome domains, outcomes, outcome measures and case-mix variables) were extracted from full-text. Third, additional references were randomly selected and reviewed in sets of 50 until saturation was reached. Saturation was defined as not finding any new MS outcomes.

b) Patient focus groups

Four three-hour focus groups, ranging from four to eight participants each, were organized for people with MS who were being treated in a top-clinical teaching hospital in the center of the Netherlands. Eligibility criteria and a topic list were determined by the research team (Appendix B). Patients were invited by an invitation letter and provided written informed consent before being audio recorded. Participants (n=30) varied in terms of education level, disease duration, clinical disease course, age and sex (Table 1). Outcomes mentioned by the participants were recorded by the main researcher and validated with the participants at the end of each meeting.

c) Rand-modified Delphi process

The RAND-modified Delphi process (Fitch et al., 2001) consisted of four online consensus meetings and three online voting rounds with a multidisciplinary expert panel (Fig. 1). The panel consisted of seventeen experts (Appendix E) from eight high-prevalent MS countries and five

Table 1
Characteristics of patient focus group participants

		N=30
Age		51.2±11.6
Sex, male		43.3%
Clinical disease course		
	RRMS	53.3%
	SPMS	26.7%
	PPMS	20.0%
Disease duration		10.2±6.0
Education level		
	Low	13.3%
	Middle	36.7%
	High	50.0%

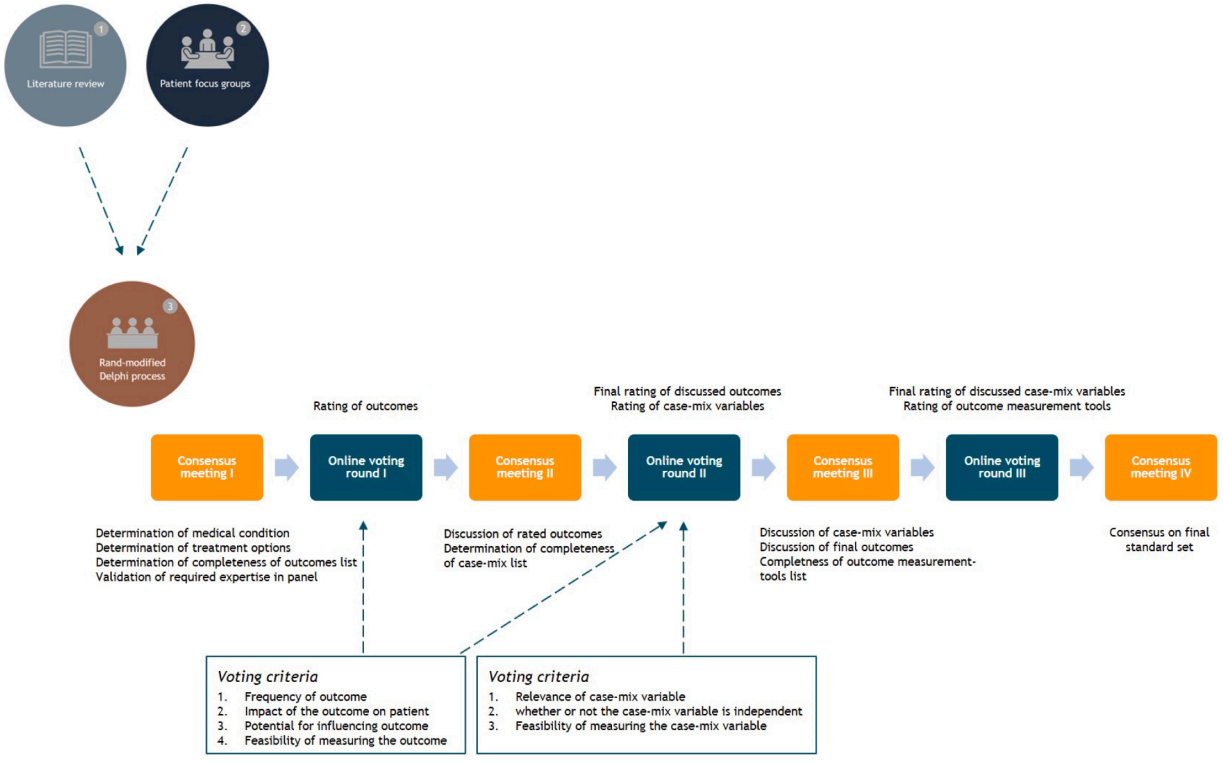


Fig. 1. Overview of the Rand-modified Delphi process

disciplines (Table 2). Experts were required to have no conflicting interests.

In the first consensus meeting, the medical condition, patient population and included treatment options were defined in detail. Furthermore, the expert panel discussed if all required expertise was present. Based on the discussion, an occupational therapist was added. Lastly, the completeness of the list of outcomes, gathered in the literature review and patient focus groups, was validated. In the first online voting round that followed, panelists rated the importance of each outcome on a 9-point Likert scale ranging from 1 (not important at all) to 9 (very important). Voting criteria were: 1). Frequency of the outcome in clinical practice, 2). Impact of the outcome on patients, 3). Potential for influencing the outcome, and 4). Feasibility of measuring the outcome. The 'potential for influencing the outcome' refers to the influence of treatment on the outcome, taking into account all treatment options. Outcomes that were rated as 'very important' (7-9 points) by more than 70% of the expert panel were included in the set. Outcomes that were rated as very important by 50-70% of the panel were up for discussion. Outcomes that were rated as very important by less than 50% of the panel were excluded.

In the second consensus meeting, results of the first online voting round were discussed. Moreover, the expert panel validated the completeness of the total list of case-mix variables, gathered in the literature review. In the online voting round that followed, the outcomes that were still up for discussion were rated again using the same criteria from the prior round. Furthermore, the expert panel rated the case-mix variables using the following criteria: 1) Relevance of the case-mix variable, 2) whether or not the case-mix variable is independent, and 3) Feasibility of measuring the case-mix variable.

In the third consensus meeting, voting results were discussed aiming to reach unanimous consensus for the final outcomes and the case-mix variables by voting during the meeting. In addition, a list of potential outcome measures was discussed and validated. In the last online voting round, panelists selected an outcome measure for every outcome in the set.

In the final consensus meeting, voting results were discussed aiming to reach unanimous consensus on the total set of outcomes, case-mix variables, outcome measures and timing by voting during the meeting.

3. Results

3.1. List of potential outcomes and case-mix variables

The literature review produced 15,328 articles. Saturation was reached after screening 200 + 50 references (Appendix A). Forty-nine potential outcomes (Appendix C) and 30 potential case-mix variables (Appendix D) were identified in the literature review.

In the patient focus group 32 potential outcomes were recognized. Three identified outcomes were not previously found in the literature review, namely: feeling cold, sleep quality, and drowsy feeling (side effect). These were not experienced by all participants. Overall, patients indicated that outcomes in the symptom, functional status and quality of life domain were most important. Furthermore, patients emphasized the importance of their own disease experience versus objective clinical indicators (e.g. lesions on MRI images). In addition, they expressed that some outcome measurements felt like a snapshot of their health at a specific day, which was not always representative for a longer period.

Table 2
Characteristics of expert panelists

		N=17
Age		47.4±10.2
Sex, male		52.9%
Country		
	The Netherlands	4
	United States of America	4
	Portugal	3
	Ireland	1
	India	1
	New Zealand	2
	Switzerland	1
	Turkey	1
Disciplines		
	Physiotherapists	6
	Neurologists	4
	MS nurses	3
	Patient representatives	2
	Health economist	1
	Occupational therapist	1

3.2. Patient population

The standard set of outcomes for MS is applicable to all adults (≥ 18 years) diagnosed with MS, taking into account all treatment options (medication, non-medication). The set can be used for measuring both newly diagnosed patients and for patients those already under treatment. The expert panel decided to use the recently redefined MS classification of Lublin et al. (Lublin, 2014) instead of the original McDonald classification (McDonald et al., 2001). Outcomes were considered equally important for all people with MS, though severity, frequency and recovery of symptoms differ between disease courses. The expert panel underlined that there should be special attention for progressive patients in the standard outcome set since this patient group is often overlooked.

3.3. Standard set of outcomes

The attendance rate of the consensus meetings was at least 65% for every meeting, and the response rate for the voting rounds was at least 82% for all rounds. Unanimous consensus was reached after four 1½-hour consensus meetings and three online voting rounds. The final set consists of fourteen outcomes divided over four domains (Fig. 2). Table 3 reports on the measuring details for the implementation of the standard outcome set for MS.

Domain 1: Disease activity [Outcomes 1-3]

The expert panel selected the 'number of new and/or active lesions per year (T2/T1 MRI)' [1], 'number of relapses per year' [2], and the 'transition from relapsing course to progressive course' [3] as outcomes in the domain 'disease activity'. The outcomes related to relapses [2 and 3] are primarily relevant to patients with a relapsing disease course.

Evidence of enhanced lesions on the T1 + gadolinium MRI and new lesions on the T2 follow-up MRI represent disease activity and are important reasons for switching treatments. However, T1 + gadolinium imaging only shows the enhanced active lesions of the last few months, while T2 imaging enables measurement of total lesion load whether currently active or not. Unfortunately, frequent T1 + gadolinium imaging is not (financially) feasible for every country/institution and, according to some people in the patient focus groups, the administration of gadolinium can sometimes be experienced as burdening by patients. Since clinical trials display both the total lesion load or increased lesion load (T2) and the active inflammation (T1), the expert panel decided to incorporate both new and/or active lesions in the standard outcome set. Measuring the T2 lesion load yearly is required according to the international protocol for follow-up of established MS (Traboulsee et al., 2016). Measuring T1 enhanced active lesions with gadolinium is considered optional.

The 'number of relapses per year' [2] was selected as an important indicator of disease activity. Relapse dates should be recorded so that the annual relapse rate can be calculated and the pattern of relapses can be determined.

'Transition from a relapsing disease course to a progressive disease course' [3] was selected since a progressive disease course leads to increased long-term disability. The transition from a relapsing course to a progressive course, as specified in the Lublin classifications (Lublin,

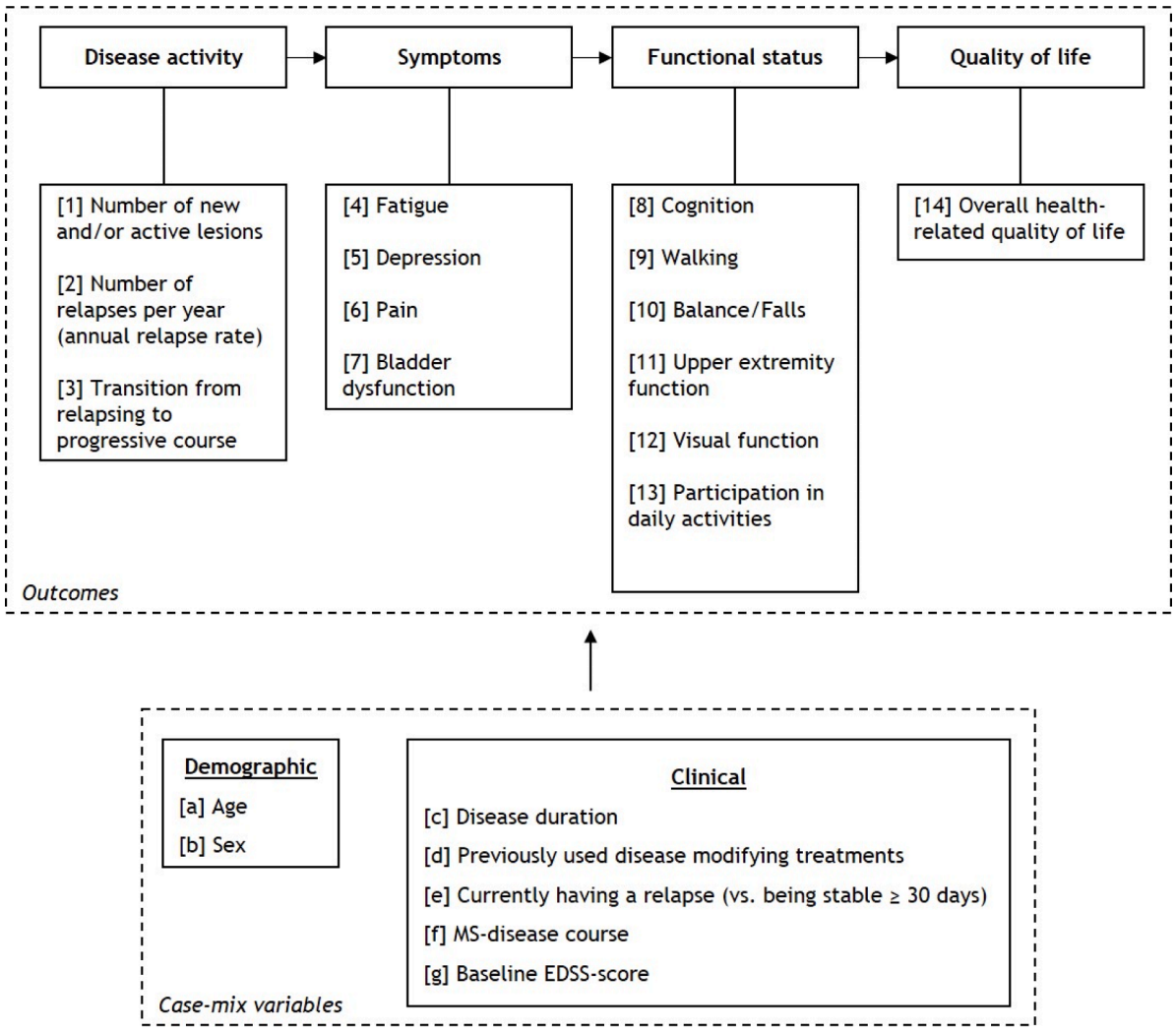


Fig. 2. Overview of the standard outcome set for MS based on the framework of Valderas and Alonso (Valderas and Alonso, 2008)

2014), can be established by a certified MS expert.

Domain 2: Symptoms [Outcomes 4-7]

The expert panel selected 'fatigue' [4], 'depression' [5], 'pain' [6] and 'bladder dysfunction' [7] as outcomes in the 'symptoms' domain.

'Fatigue' [4] was selected because it is experienced by many patients and is considered very disabling. As an outcome measure, the 5-item Modified Fatigue Impact Scale (MFIS-5) (Ritvo et al., 1997) was selected because of its availability, low administrative burden for patients, and good psychometric properties (Meca-Lallana et al., 2019).

'Depression' [5] was identified as an important outcome because it has a high impact on daily life and is a sensitive and difficult topic to talk about. The panel selected the 9-item Patient Health Questionnaire (PHQ-9) as outcome measure for depression because of its free availability, easy administration and high sensitivity (84%) (Marrie et al., 2018).

'Pain' [6] was selected because it is a treatable and very important outcome for patients. The expert panel selected the visual analog scale (VAS) for pain as outcome measure because of its low administrative burden for patients.

Furthermore, 'bladder dysfunction' [7] was selected since this symptom can be (socially) disabling. Indication of a bladder dysfunction can be detected by the single item #20 in the Multiple Sclerosis Impact Scale (MSIS-29) (Hobart et al., 2001). Adding a detailed questionnaire or physical test on bladder dysfunction to the set was not considered feasible for every MS center. Furthermore, the panel advised that patients should be referred to a urologist for clinical examination in case a bladder dysfunction is indicated.

Domain 3: Functional status [Outcomes 8-13]

In the 'functional status' domain, 'cognition' [8], 'walking' [9], 'balance/falls' [10], 'upper extremity function' [11], 'visual function' [12] and 'participation in daily activities' [13] were selected as the most relevant outcomes.

'Cognition'[8] was selected because of its impact on daily life and patients' fear of losing their cognitive functions. The Symbol Digit Modalities Test (SDMT) was selected as the appropriate outcome measure. The SDMT focuses on cognitive processing speed (CPS) and is commonly used because of its ease of administration, short duration (5min), and good methodological properties (Benedict et al., 2017).

The expert panel selected both 'walking' [9] and 'balance' [10]. They discussed the overlap with 'mobility' since it includes both walking and balance/falls, and could encompass ability to move in bed, stand up or move from a bed to a chair. Ultimately, mobility was excluded because the expert panel preferred single outcomes instead of composite outcomes. For walking, the Timed 25-Foot Walk (T25FW) test was chosen as outcome measure. The T25FW has good psychometric properties and can be performed with and without the use of a walking aid (Fischer et al., 1999, Larson et al., 2013). Although the T25FW test is an imperial measure (equivalent: 6.7 meters), the test is frequently used in both North-America and Europe. The expert panel emphasized the importance of a notification of the patient's used walking aid (Kieseier and

Table 3
Details of the standard outcome set for MS

Domain	Outcome	Details	Timing
Disease activity	[1] Number of new and/or active lesions *	MRI T1 and/or MRI T2 imaging	Annually
	[2] Number of relapses per year (Annual Relapse Rate) °	Date of relapse	Ongoing
	[3] Switch from relapsing to progressive ° ǂ	Date of switch	Ongoing
Symptoms	[4] Fatigue	5-item Modified Fatigue Impact Scale (MFIS-5)	6-monthly
	[5] Depression	9-item Patient Health Questionnaire (PHQ-9)	6-monthly
	[6] Pain	Visual Analog Scale (VAS)	6-monthly
	[7] Bladder dysfunction	Multiple Sclerosis Impact Scale (MSIS-29) ∞	6-monthly
Functional status	[8] Cognition	Symbol Digit Modalities Test (SDMT)	6-monthly
	[9] Walking	Timed 25-foot walk test (T25FW) ~	6-monthly
	[10] Balance/Falls	Timed up and go (TUG), and number of falls	6-monthly
	[11] Upper extremity function	Nine-Hole Peg Test (NHPT)	6-monthly
	[12] Visual function	Low-contrast letter acuity test (LCLA)	6-monthly
	[13] Participation in daily activities	Multiple Sclerosis Impact Scale (MSIS-29) §	6-monthly
Quality of Life	[14] Overall health-related quality of life	Multiple Sclerosis Impact Scale (MSIS-29) ¶	6-monthly

* measuring T1 enhanced active lesion is considered optional
° applicable for relapsing patients
ǂ clinical course according to (Lublin et al. 2014)
∞ indication of bladder dysfunction by single item #20 of MSIS-29 (Hobart et al., 2001)
~ with additional note of Disease Steps (0 = Normal; 1 = Mild disability, mild symptoms or signs; 2 = Moderate disability, visible abnormality of gait; 3 = Early cane, intermittent use of cane; 4 = Late cane, cane-dependent; 5 = Bilateral support; 6 = Confined to wheelchair; and U = Unclassifiable) (Kieseier and Pozzilli, 2012)
§ indication of participation level in daily functioning by single items #12-19 of MSIS-29 (Hobart et al., 2001)
¶ total score of MSIS-29, consisting of both the physical and psychological subscale (Hobart et al., 2001)

Table 4
Case-mix variables in the standard set

Domain	Case-mix variable	Details	Timing
Demographic	[a] Age	Date of birth	At baseline – diagnosis
	[b] Sex	Sex at birth	At baseline - diagnosis
Clinical	[c] Disease duration	Date diagnosis	At baseline - diagnosis
	[d] Previously used disease modifying treatments (DMTs)	a) Type DMT, b) Date start DMT	Ongoing
	[e] Currently having a relapse (vs being stable for 30 days)	Date of relapse	Ongoing
	[f] Type of MS	Type of MS at diagnosis ǂ	At baseline - diagnosis
	[g] Baseline EDSS score	Extended Disability Status Scale §	At baseline - diagnosis

ǂ clinical course according to Lublin et al. (Lublin, 2014)
§ EDSS (Kurtzke, 1983)

Pozzilli, 2012).

'Balance' [10] was selected because it is frequently affected by MS, and fear of falling among patients is high. The Timed Up and Go (TUG) test was selected as the outcome measure for balance because of its psychometric properties and its easy administration to all patients in almost every setting (Allali et al., 2012, Christopher et al., 2019). The panel also recommended recording the number of falls (per year) because a recent study showed that the TUG does not adequately discriminate 'fallers' from 'non-fallers' (Quinn et al., 2019).

'Upper extremity function' [11] was identified as very important because of its impact in daily life. Moreover, when confined to a wheelchair, treatment can still be targeted at minimizing impairment in upper extremity functions. The Nine-Hole-Peg-Test (9-HPT) (Goodkin et al., 1988, Feys et al., 2017) was selected as the outcome measure for upper extremity function, because it has good psychometric properties and is easily administered.

'Visual function' [12] was chosen because of its high prevalence and disabling impact. The Low-Contrast Letter Acuity (LCLA) test (Balcer et al., 2003) was selected as the outcome measure for visual functioning because of its applicability and affordability. LCLA has the ability to detect clinically relevant changes in visual function and has good psychometric properties (Balcer et al., 2017).

'Participation in daily activities' [13] was selected by the expert panel because patients find this overarching outcome important. The expert panel recognized that participation level can be highly impacted by outcomes identified in the 'symptoms' domain. An indication of 'participation in daily activities' can be provided by the Multiple Sclerosis Impact Scale MSIS-29, items #12-19) (Hobart et al., 2001).

Domain 4: Quality of Life [Outcome 14]

In the 'quality of life' domain, 'overall health-related quality of life' (HRQOL) [14] was selected because it reflects a patient's general, physical and emotional wellbeing. Patients indicated that this was the most important outcome for them. The MSIS-29 (Hobart et al., 2001), a patient-reported outcome measurement (PROM) specific for MS patients, was chosen because of its good psychometric properties (McGuigan and Hutchinson, 2004) and its relatively low administrative burden for patients. The MSIS-29 inventories physical and psychological wellbeing independently.

3.4. Case-mix variables

The expert panel reached consensus on a minimum set of 7 case-mix variables (Table 4). Demographic factors that were selected were 'age' [a] and 'sex' [b]. 'Disease duration' [c], 'previously used DMTs' [d], 'currently having a relapse (vs stable for 30 days)' [e], 'MS disease course (relapsing vs progressive)' [f] and 'baseline EDSS score' [g] were selected as clinical factors. The last three case-mix variables [e-g] were immediately included based on the voting results. 'Disease duration' [c] was considered relevant because of its effect on the outcomes and its use in scientific research. 'Previously used DMTs' [d] was considered important because of its frequent use in scientific research and its influence on what treatment options are still available for a patient.

4. Discussion

In this study, an international standard set of patient-relevant outcomes was developed for MS. The standard set consists of fourteen outcome measures, divided over the domains 'disease activity',

'symptoms', 'functional status' and 'quality of life', and seven case-mix variables. Outcomes not included in this set were rejected for a variety of reasons including not being relevant to the majority of patients, lack of impact, feasibility of measurement and overlapping with other outcomes.

The expert panel intentionally selected PROMs for some of the outcomes, especially in the 'quality of life' and 'symptoms' domains, because of patients' desire to measure their own disease experience versus objective clinical indicators. The expert panel deliberately selected a MS-specific HRQOL PROM instead of a general HRQOL PROM because of its clinical details and face validity. People with MS might recognize their own symptoms more easily in the questions and be more willing to fill out the questionnaire. Furthermore, the MSIS-29 measures not only overall HRQOL, but inquires about MS-specific outcomes such as bladder dysfunction, participation in daily activities, fatigue, and depression. Monitoring multiple symptoms using just one questionnaire will limit the administrative burden for patients.

Some of the outcome measures in this standard set are part of composite outcome measures such as the MS Functional Composite (MSFC) (Fischer et al., 1999) and the Expanded Disability Status Scale (Kurtzke, 1983). Although discussed because of its widely-use, the panel ultimately decided not to use overall test scores such as the MSFC or EDSS because of potential loss of information due to their overarching character and EDSS' limitations in scaling properties. Thus, the present set includes some of the components of the MSFC but not all; the T25FWT and the 9-HPT were selected, but the expert panel preferred the SDMT over the Paced Auditory Serial Addition Test (PASAT) because of its validity and reliability as a single outcome measure for cognition (Sonder et al., 2014).

Previous attempts to standardize MS outcomes are consistent with some elements of the present standard set. For instance, the outcome measures that our expert panel selected in the 'functional status' domain correspond to some of the outcome measures selected in the sets for exercise (Paul et al., 2014) and rehabilitation programs (Cohen et al., 2015, Potter et al., 2014). However, in contrast to previous attempts to standardize MS outcomes (Paul et al., 2014, Cohen et al., 2015, Potter et al., 2014, Swart et al., 2019, van Munster and Uitdehaag, 2017), the present set has a holistic character and includes multiple domains while still minimizing the number of outcomes. Furthermore, as opposed to some of the previous initiatives (Cohen et al., 2015, Potter et al., 2014, van Munster and Uitdehaag, 2017), we actively involved patients and patient representatives in the development of the set. Because of the multidisciplinary and international character of our expert panel, we expect that the present set can be valuable for and administered by MS-care providers of all disciplines and countries.

Contrary to standard sets for other chronic or neurological conditions (Verberne et al., 2019, Kampstra et al., 2019, Salinas et al., 2016), we have not selected the outcome 'survival'. Patients and panelists, including patient representatives, stated that survival, although frequently included in standard outcome sets for other medical conditions, was not specifically important to MS patients since MS is a chronic disease and mortality is often not directly related to MS.

Although the S.O.S.MS project succeeded in reaching its aim, there are some limitations to consider. First, there might have been some selection bias due to the patient focus group composition. However, we consider its influence on the final set limited since most outcomes identified in the patient focus group correspond to the results of the literature review. Additionally, patient representatives who were part of the international expert panel validated the final outcome set. Second, since we have included experts from areas where MS is most prevalent, most panelists are from high-income countries. It is therefore unknown whether it is feasible to implement the present set in low- and middle income countries. To increase generalizability the expert panel only selected outcome measures that are freely available and easy to administer in every setting.

It is important to note that standard sets are not fixed but change over time. Periodically, this set should be updated according to the latest treatment options, diagnostic criteria and available outcome measures. Furthermore, implementing the present set in daily clinical practice provides the opportunity to validate and refine the standard set as well as test its feasibility in different settings (Daeter et al., 2018).

5. Conclusion

The standard outcome set for MS developed in this study provides a guideline for measuring outcomes of MS in clinical practice and research. We expect that the present standard set can be implemented in all MS care settings due to the limited number of outcomes and the selection of freely available and easily administered outcome measures. In addition to this standard set, MS-healthcare providers can add outcomes that they deem relevant. Using this standard outcome set to monitor and (inter)nationally benchmark real-world outcomes of MS can provide the opportunity to improve patient value and ultimately support the transition towards value-based MS care.

CRediT authorship contribution statement

K. Daniels: Conceptualization, Methodology, Formal analysis, Writing – original draft. **S.T.F.M. Frequin:** Conceptualization, Resources, Writing – review & editing. **E.M.W. van de Garde:** Writing – review & editing. **D.H. Biesma:** Writing – review & editing. **P.J. van der Wees:** Conceptualization, Methodology, Writing – review & editing. **P. B. van der Nat:** Conceptualization, Methodology, Writing – original draft. **Burcu Ersoz Huseyinsinoglu:** Writing – review & editing. **Aliza Bitton Ben-Zacharia:** Writing – review & editing. **E.T. Cohen:** Writing – review & editing. **Paulo Jorge Correia Gonçalves:** Writing – review & editing. **J. Jolijn Kragt:** Writing – review & editing. **Sinéad M. Hynes:** Writing – review & editing. **Frances Elizabeth Marron:** Writing – review & editing.

Declaration of Competing Interest

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

KD, PBVDN, STFMF, PJVDW and EMWVDG received funding for a VBHC investigator-institution initiated research for MS from Roche Nederland B.V. Funding paid to Institution. Roche Nederland B.V. did not have any influence on the present study's design, execution, authorship or publication.

Ethical approval

The protocol of this study was approved by the Medical research Ethics Committees United (MEC-U) (registration number: W19.159) and by the institutional review board of St. Antonius Hospital (registration number: R&D/Z19.077).

Acknowledgements

We would like to express our gratitude towards all patients who have participated in the patient focus groups. Furthermore, we would like to thank all MS experts who have contributed to the Rand-modified Delphi process.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.msard.2022.104461.

References

Allali, G, Laidet, M, Assal, F, Beauchet, O, Chofflon, M, Armand, S, et al., 2012. Adapted timed up and go: A rapid clinical test to assess gait and cognition in multiple sclerosis. Amato Benedetti, Cattaneo, Chiaravalloti, Feuillet, Hauser, Herman, Linden, Nilsagard, Perret, Pettersson, Podsiadlo, Polman, Tanji, Tombaugh, Wechsler, Yardley, Zigmond, Zimmermann B, editor Eur. Neurol. [Internet]. 67 (2), 116–120. Available from. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc9&NEWS=N&AN=2012-07088-009.

Balcer, LJ, Baier, ML, Cohen, JA, Kooijmans, MF, Sandrock, AW, Nano-Schiavi, ML, et al., 2003. Contrast letter acuity as a visual component for the Multiple Sclerosis Functional Composite. Neurology 61 (10), 1367–1373. Nov.

Balcer, LJ, Raynowska, J, Nolan, R, Galetta, SL, Kapoor, R, Benedict, R, et al., 2017. Validity of low-contrast letter acuity as a visual performance outcome measure for multiple sclerosis. Mult. Scler. 23 (5), 734–747. Apr.

Benedict, RH, DeLuca, J, Phillips, G, LaRocca, N, Hudson, LD, Rudick, R, 2017. Multiple Sclerosis Outcome Assessments Consortium. Validity of the Symbol Digit Modalities Test as a cognition performance outcome measure for multiple sclerosis. Mult. Scler. 23 (5), 721–733. https://doi.org/10.1177/1352458517690821.

Christopher, A, Kraft, E, Olenick, H, Kiesling, R, Doty, A., 2019. The reliability and validity of the Timed Up and Go as a clinical tool in individuals with and without disabilities across a lifespan: a systematic review. Disabil. Rehabil. 1–15. Oct.

Cohen, ET, Potter, K, Allen, DD, Bennett, SE, Brandfass, KG, Widener, GL, et al., 2015. Selecting Rehabilitation Outcome Measures for People with Multiple Sclerosis. Int. J. MS Care 17 (4), 181–189.

Daeter, EJ, Timmermans, MJC, Hirsch, A, Lipsic, E, Houterman, S, Amoroso, G, et al., 2018. Defining and Measuring a Standard Set of Patient-Relevant Outcomes in Coronary Artery Disease. Am. J. Cardiol. 121 (12), 1477–1488.

de Roos, P, Bloem, BR, Kelley, TA, Antonini, A, Dodel, R, Hagell, P, et al., 2017. A Consensus Set of Outcomes for Parkinson's Disease from the International Consortium for Health Outcomes Measurement. J. Parkinsons Dis. 7 (3), 533–543.

Feys, P, Lamers, I, Francis, G, Benedict, R, Phillips, G, LaRocca, N, Hudson, LD, Rudick, R, 2017. Multiple Sclerosis Outcome Assessments Consortium. The Nine-Hole Peg Test as a manual dexterity performance measure for multiple sclerosis. Mult. Scler. 23 (5), 711–720. https://doi.org/10.1177/1352458517690824.

Fischer, JS, Rudick, RA, Cutter, GR, Reingold, SC., 1999. The Multiple Sclerosis Functional Composite Measure (MSFC): an integrated approach to MS clinical outcome assessment. National MS Society Clinical Outcomes Assessment Task Force. Mult. Scler. 5 (4), 244–250. Aug.

Fitch, K, Bernstein, SJ, Mcdonnell, J, Kahan, JP., 2001. The RAND /UCLA Appropriateness Method User ' s Manual Appropriateness Method User ' s Manual Approved for Public Release Approved for Public Release. Transformation 109 p.

Goodkin, DE, Hertsgaard, D, Seminary, J., 1988. Upper extremity function in multiple sclerosis: improving assessment sensitivity with box-and-block and nine-hole peg tests. Arch. Phys. Med. Rehabil. 69 (10), 850–854. Oct.

Hobart, J, Lamping, D, Fitzpatrick, R, Riazi, A, Thompson, A., 2001. The multiple sclerosis impact scale (MSIS-29) a new patient-based outcome measure. Brain 124 (5), 962–973.

Kampstra, NA, Grutters, JC, Van Beek, FT, Culver, DA, Baughman, RP, Renzoni, EA, et al., 2019. First patient-centred set of outcomes for pulmonary sarcoidosis: A multicentre initiative. BMJ Open Respir. Res. 6 (1), 1–8.

Kieseier, BC, Pozzilli, C, 2012. Assessing walking disability in multiple sclerosis. Mult. Scler. 18 (7), 914–924. https://doi.org/10.1177/1352458512444498.

Kurtzke, JF., 1983. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). Neurology 33 (11), 1444–1452. Nov.

Larson, RD, Larson, DJ, Baumgartner, TB, White, LJ., 2013. Repeatability of the timed 25-foot walk test for individuals with multiple sclerosis. Clin. Rehabil. 27 (8), 719–723. Aug.

Lublin, FD., 2014. New multiple sclerosis phenotypic classification. Hawker Lublin, Polman. Wolinsky L, editor Eur. Neurol. [Internet] 72 (Supp1), 1–5. Available from. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc11&NEWS=N&AN=2014-41521-002.

Marrie, RA, Zhang, L, Lix, LM, Graff, LA, Walker, JR, Fisk, JD, et al., 2018. The validity and reliability of screening measures for depression and anxiety disorders in multiple sclerosis. Mult. Scler. Relat. Disord. 20, 9–15. November 2017.

Marziniak, M, Ghorab, K, Kozubski, W, Pfleger, C, Sousa, L, Vernon, K, et al., 2016. Variations in multiple sclerosis practice within Europe - Is it time for a new treatment guideline? Mult. Scler. Relat. Disord. 8 (2016), 35–44.

McDonald, WI, Compston, A, Edan, G, Goodkin, D, Hartung, H, Lublin, FD, et al., 2001. Recommended diagnostic criteria for multiple sclerosis: guidelines from the International Panel on the diagnosis of multiple sclerosis. Ann. Neurol. Off J. Am. Neurol. Assoc. Child. Neurol. Soc. 50 (1), 121–127.

McGuigan, C, Hutchinson, M., 2004. The multiple sclerosis impact scale (MSIS-29) is a reliable and sensitive measure. J. Neurol. Neur. Psyc. 75 (2), 266–269. Feb.

Meca-Lallana, V, Brañas-Pampillón, M, Higueras, Y, Candeliere-Merlicco, A, Aladro-Benito, Y, Rodríguez-De la Fuente, O, et al., 2019. Assessing fatigue in multiple sclerosis: Psychometric properties of the five-item Modified Fatigue Impact Scale (MFIS-5). Mult. Scler. J. Exp. Transl. Clin. 5 (4), 205521731988798.

Paul, L, Coote, S, Crosbie, J, Dixon, D, Hale, L, Holloway, E, McCrone, P, Miller, L, Saxton, J, Sincock, C, White, L, 2014 Oct. Core outcome measures for exercise studies in people with multiple sclerosis: recommendations from a multidisciplinary consensus meeting. Mult. Scler 20 (12), 1641–1650. https://doi.org/10.1177/1352458514526944.

Porter, ME., 2010. What is value in health care? N. Engl. J. Med. 363 (26), 2477–2481. Dec.

Potter, K, Cohen, ET, Allen, DD, Bennett, SE, Brandfass, KG, Widener, GL, et al., 2014. Outcome measures for individuals with multiple sclerosis: recommendations from the American Physical Therapy Association Neurology Section task force. Phys. Ther. 94 (5), 593–608. May.

Quinn, G, Comber, L, McGuigan, C, Galvin, R, Coote, S., 2019. Discriminative ability and clinical utility of the Timed Up and Go (TUG) in identifying falls risk in people with multiple sclerosis: a prospective cohort study. Clin. Rehabil. 33 (2), 317–326. Feb.

Ritvo, PG, Fischer, JS, Miller, DM, Andrews, H, Paty, DW, LaRocca, NG., 1997. Multiple sclerosis quality of life inventory: a user's manual. New York Natl. Mult. Scler. Soc 65.

Salinas, J, Sprinkhuizen, SM, Ackerson, T, Bernhardt, J, Davie, C, George, MG, et al., 2016. An international standard set of patient-centered outcome measures after stroke. Stroke 47 (1), 180–186.

Sonder, JM, Burggraaff, J, Knol, DL, Polman, CH, Uitdehaag, BM, 2014. Comparing long-term results of PASAT and SDMT scores in relation to neuropsychological testing in multiple sclerosis. Mult. Scler. 20 (4), 481–488. https://doi.org/10.1177/1352458513501570.

Swart, ECS, Neilson, LM, Good, CB, Shrank, WH, Henderson, R, Manolis, C, Parekh, N, 2019. Determination of Multiple Sclerosis Indicators for Value-Based Contracting Using the Delphi Method. J. Manag. Care Spec. Pharm. 25 (7), 753–760. https://doi.org/10.18553/jmcp.2019.25.7.753.

Traboulsee, A., Simon, J.H., Stone, L., Fisher, E., Jones, D.E., Malhotra, A., Newsome, S. D., Oh, J., Reich, D.S., Richert, N., Rammohan, K., Khan, O., Radue, E.-W., Ford, C., Halper, J., 2016. Revised Recommendations of the Consortium of MS Centers Task Force for a Standardized MRI Protocol and Clinical Guidelines for the Diagnosis and Follow-Up of Multiple Sclerosis. AJNR Am. J. Neuroradiol. 37 (3), 394–401.

Valderas, JM, Alonso, J., 2008. Patient reported outcome measures: A model-based classification system for research and clinical practice. Qual. Life Res. 17 (9), 1125–1135.

van Munster, CE, Uitdehaag, BM, 2017 Mar. Outcome Measures in Clinical Trials for Multiple Sclerosis. CNS Drugs 31 (3), 217–236. https://doi.org/10.1007/s40263-017-0412-5.

Verberne, WR, Das-Gupta, Z, Allegretti, AS, Bart, HAJ, van Biesen, W, García-García, G, et al., 2019. Development of an International Standard Set of Value-Based Outcome Measures for Patients With Chronic Kidney Disease: A Report of the International Consortium for Health Outcomes Measurement (ICHOM) CKD Working Group. Am. J. Kidney Dis. 73 (3), 372–384.