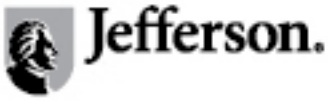

2-21-2019

Development and Validation of the FSIQ-RMS: A New Patient-Reported Questionnaire to Assess Symptoms and Impacts of Fatigue in Relapsing Multiple Sclerosis

Stacie Hudgens
Clinical Outcomes Solutions

René Schüler
Actelion

Jonathan Stokes
Adelphi Values

Sonya Eremenco
Evidera

Elke Hunsche
Actelion

See next page for additional authors

Authors
Stacie Hudgens, René Schüler, Jonathan Stokes, Sonya Eremenco, Elke Hunsche, and Thomas P. Leist

VALUE IN HEALTH ■ (2019) ■–■

Available online at www.sciencedirect.com

ScienceDirect

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

Development and Validation of the FSIQ-RMS: A New Patient-Reported Questionnaire to Assess Symptoms and Impacts of Fatigue in Relapsing Multiple Sclerosis

Stacie Hudgens, MA (AbD) [1,*], *René Schüler, MSc* [2], *Jonathan Stokes, MBA* [3], *Sonya Eremenco, MA* [4], *Elke Hunsche, PhD* [5], *Thomas P. Leist, MD, PhD* [6]

[1]*Clinical Outcomes Solutions, Tucson, AZ, USA;* [2]*Global Market Access & Pricing, Actelion, Singapore;* [3]*Patient-Centered Outcomes, Adelphi Values, Boston, MA, USA;* [4]*Outcomes Research, Evidera, Bethesda, MD, USA;* [5]*Global Market Access & Pricing, Actelion, Allschwil, Switzerland;* [6]*Comprehensive Multiple Sclerosis Center, Thomas Jefferson University, Philadelphia, PA, USA*

ABSTRACT

Objectives: A new patient-reported outcome (PRO) instrument to measure fatigue symptoms and impacts in relapsing multiple sclerosis (RMS) was developed in a qualitative stage, followed by psychometric validation and migration from paper to an electronic format. **Methods:** Adult patients with relapsing-remitting multiple sclerosis (RRMS) were interviewed to elicit fatigue-related symptoms and impacts. A draft questionnaire was debriefed in cognitive interviews with further RRMS patients, and revised. Content confirmation interviews were conducted with patients with progressive-relapsing multiple sclerosis (PRMS) and relapsing secondary-progressive multiple sclerosis (RSPMS). Psychometric analyses used data from adult patients with different RMS subtypes and matched non-RMS controls in a multicenter, observational study. After item reduction, the final instrument was migrated to a smartphone (eDiary) and usability was confirmed in interviews with additional adult RMS patients. **Results:** The qualitative stage included 37 RRMS, 5 PRMS, and 5 RSPMS patients. Saturation of concepts was reached during concept elicitation. Cognitive interviews confirmed that participants understood the instructions, items, and response options of the instrument—named FSIQ-RMS—as intended. Psychometric validation included 164 RMS and 74 control patients. Internal consistency and test–retest reliability were demonstrated. The symptoms domain discriminated along the RMS symptom-severity continuum and between patients and controls. Patients were able to attribute fatigue-related symptoms to RMS. Usability and conceptual equivalence of the eDiary were confirmed (n = 10 participants). **Conclusions:** With 7 symptom items and 13 impact items (in 3 impacts subdomains: physical, cognitive and emotional, and coping) after item reduction, the FSIQ-RMS is a comprehensive, valid, and reliable measure of fatigue-related symptoms and impacts in RMS patients.

Keywords: fatigue, patient-reported outcome, relapsing multiple sclerosis, symptoms

Introduction

Fatigue is one of the most common symptoms of multiple sclerosis (MS),[1–3] a chronic autoimmune disorder of the central nervous system.[4] Often described as a feeling of extreme mental or physical exhaustion, MS-related fatigue has far-reaching effects on quality of life, employment, and productivity,[5–7] imposing limitations independently of MS-related physical disability and depression.[8]

Given the subjective nature of fatigue, the effect of treatment on fatigue in MS is best assessed via a patient-reported outcome (PRO) instrument.[9] Such a tool would also be valuable in clinical practice, as MS guidelines and proposed quality-of-care measures call for fatigue assessment using a PRO instrument with demonstrated consistency, reliability, and validity.[10,11] The 2009 Food and Drug Administration (FDA) PRO guidance, the FDA's 2015 roadmap to patient-focused outcome measurement in clinical trials, and the 2015 recommendations from the ISPOR Clinical Outcomes Assessment—Emerging Good Practices for Outcomes Research Task Force all emphasize the need to document the content and construct validity, reliability, and sensitivity to change of a PRO instrument.[12–14]

Although available PRO instruments have been used to measure fatigue in MS patients,[10,15–20] review of their

* *Address correspondence to:* Stacie Hudgens, Clinical Outcomes Solutions, 1790 East River Road, Suite #205, Tucson, AZ 85718, USA.
E-mail: stacie.hudgens@clinoutsolutions.com

1098-3015 - see front matter
https://doi.org/10.1016/j.jval.2018.11.007

measurement properties suggests shortcomings in terms of current standards for PRO instrument development. For instance, the 9-item Fatigue Severity Scale (FSS) and the 21-item Modified Fatigue Impact Scale (MFIS) do not fit the assumption of unidimensionality, and so studies using their global scores may need to be reevaluated.[21,22]

A primary requirement is to develop and validate a PRO instrument in the disease-specific target patient population for intended use. The clinical course of MS broadly falls into relapsing and progressive forms (RMS and PMS, respectively).[23] Three RMS subtypes are recognized.[23–25] Relapsing-remitting MS (RRMS), seen in approximately 80% to 85% of MS patients, is characterized by acute exacerbations with full or partial recovery, and patients are generally stable between exacerbations. Progressive-relapsing MS (PRMS), which afflicts roughly 5% of MS patients, is marked by progression starting at disease onset, with occasional relapses. Relapsing secondary-progressive MS (RSPMS) gradually progresses later (after a median of 20 years) in the disease course of approximately 65% to 70% of RRMS patients.

As symptoms and their impacts could differ among these various forms of MS, a PRO instrument should be validated in the specific MS subtypes for which it will be used. A unidimensional version of the FIS (U-FIS) has been developed to address the multidimensionality of the original 40-item Fatigue Impact Scale (FIS) and the MFIS,[17,26] but some participants in its validation research were recruited from patient organizations rather than clinical centers, precluding confirmation of all participants' MS diagnoses or their type of MS. Some other MS-specific fatigue questionnaires, such as the FSS,[15] were developed in patients with PMS rather than RMS, which may affect their applicability as outcome measures in RMS.

Another issue is whether a questionnaire comprehensively measures MS symptoms and their impact; for example, the FSS does not include items regarding cognitive fatigue,[27] whereas the FIS and MFIS assess the impact of symptoms on patient functioning, rather than measuring the severity of the symptoms themselves.[2,10,16]

The objective of the present research was to address limitations of existing MS-specific instruments by developing a new, content-valid, concise PRO instrument to assess fatigue symptoms relevant to patients within the spectrum of RMS and the relevant impact of these symptoms on patients' lives, in accordance with the FDA PRO guidance.

Methods

Overview

Development and psychometric validation of the new PRO instrument was conducted in 6 research stages, preceded by a literature review (Fig. 1). Development of the draft questionnaire was based on qualitative research in patients with RMS, including concept elicitation and cognitive interviews (n = 17 and 20, respectively) in patients with RRMS, and content confirmation interviews in patients with PRMS (n = 5) and RSPMS (n = 5). Finalization of the new instrument and psychometric validation analyses were based on data gathered in an observational study, a real-world cross-sectional study, and a study to confirm the conceptual equivalence and usability of an electronic version of the questionnaire.

All study-stage protocols were approved by centralized Institutional Review Boards, and all participants provided written informed consent. All interviews were conducted by experienced interviewers trained in the use of semistructured interview guides specific to each study stage.

Literature Review

Literature reviews were conducted to identify evidence on fatigue-related symptoms of RMS and their impacts. Relevant data were used to inform development of preliminary conceptual frameworks for RMS symptoms and impacts, which in turn were used to inform the concept elicitation interview guides. Searches were conducted in Medline and covered a 10-year period for RRMS-related searches (January 2000 to September 2010) and for subsequent searches pertaining to PRMS and RSPMS (January 2002 to February 2012).

Qualitative Research

Stage 1: concept elicitation interviews in RRMS

Face-to-face concept elicitation interviews in patients with RRMS were conducted to identify and describe fatigue-related symptoms and impacts from the patient perspective. The FIS was also cognitively debriefed to determine the extent to which it may assess RRMS symptoms and impacts and therefore be fit for purpose as an efficacy endpoint measure.

In these concept elicitation interviews, each participant was asked open-ended questions to spontaneously elicit fatigue-related symptoms of RRMS and their impacts as experienced by the participant. Probing questions were posed if the initial open-ended queries did not elicit an explicit description of a fatigue-related symptom or impact.

After qualitative analysis of the interview transcripts, concept saturation was assessed. Saturation was considered to be achieved at the point at which no new, relevant concepts emerged from subsequent interviews. A sample size of 15 was anticipated to be sufficient to achieve saturation.[28] Two additional participants were recruited to account for no-shows or cancellations.

As described in the Results, the FIS would have required too many modifications to attain content validity in an RRMS population. Therefore, a new PRO instrument to assess fatigue-related symptoms of RRMS and their impacts was drafted by the investigators based on the concepts spontaneously elicited in interviews. After the preliminary conceptual framework, the instrument comprised a symptoms domain with a 24-hour recall period and an impacts domain with a 7-day recall period. Concepts reported by more than 5 (>29.4%) of the 17 participants were selected for inclusion in the initial paper questionnaire. The new instrument was named the "Fatigue Symptoms and Impacts Questionnaire—Relapsing Remitting Multiple Sclerosis" (FSIQ-RRMS).

Stage 2: cognitive interviews in RRMS

Face-to-face cognitive interviews were conducted to assess the relevance, comprehensibility, acceptability, and comprehensiveness of its items, in addition to the interpretability and appropriateness of instructions, response options, and the recall period of the FSIQ-RRMS. These cognitive interviews included 20 patients with RRMS who had not been included in stage 1.

An item was subject to revision if at least 5 (≥25.0%) of the 20 respondents gave consistent recommendations for its change (eg, rewording or deletion), demonstrated difficulty interpreting it, or did not interpret it as intended. The same threshold was applied for the addition of suggested new items to the questionnaire.

Stage 3: content confirmation interviews in PRMS and RSPMS

After stage 2, it was decided to expand the target population for the new PRO instrument to encompass a broader range of RMS patients. One-on-one content confirmation interviews were conducted in patients with PRMS or RSPMS to test the understandability, comprehensiveness, and relevance of the instrument in these RMS subtypes.

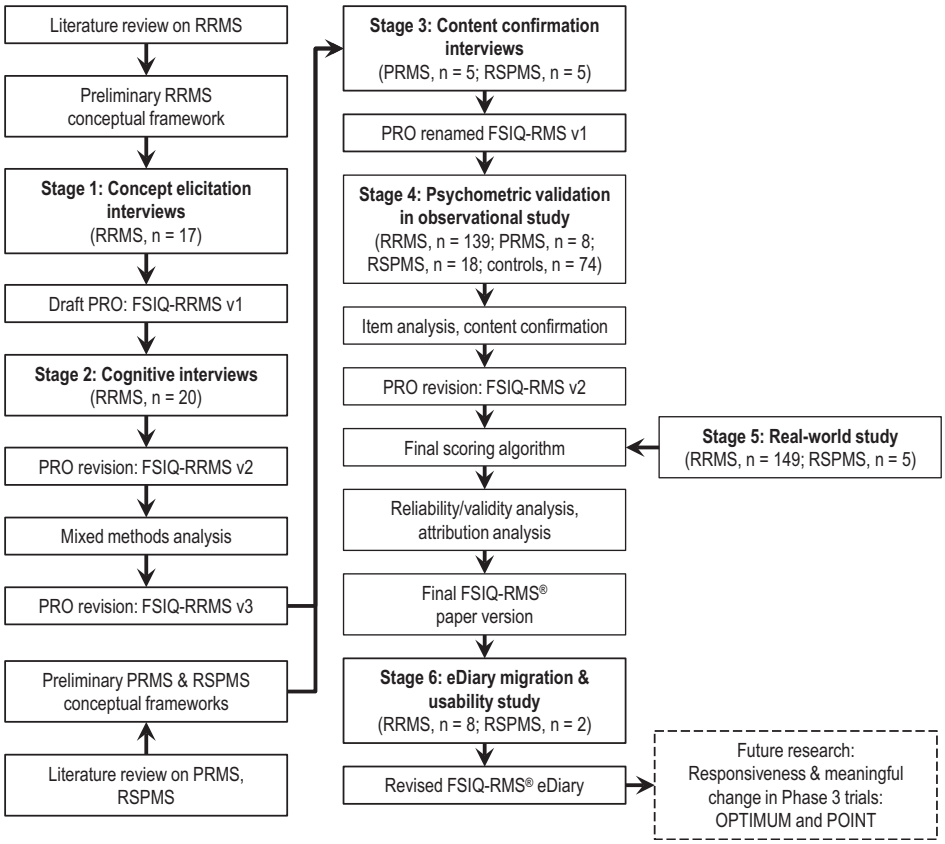

Fig. 1 – Study flow. MS, multiple sclerosis; PRO, patient-reported outcome; PRMS, progressive-relapsing MS; RRMS, relapsing-remitting MS; RSPMS, relapsing secondary-progressive MS.

Concept elicitation and cognitive interviews were conducted using semistructured interview guides based on preliminary conceptual frameworks developed for RSPMS and PRMS fatigue-related symptoms and impacts. Interview findings indicated that no revisions were required for application of the PRO instrument to patients with PRMS and RSPMS. The instrument was renamed the "Fatigue Symptoms and Impacts Questionnaire—Relapsing Multiple Sclerosis" (FSIQ-RMS).

Psychometric Validation

Stage 4: observational study

Psychometric validation was performed in an observational study that also included a control group of non-MS patients to demonstrate the ability of RMS patients to attribute fatigue symptoms to their disease (n = 164 patients with RMS + 74 controls).

Psychometric analyses included tests of item response and dimensionality, Rasch and exploratory factor analyses to assess content validity, and analyses of construct validity, internal consistency reliability, and test–retest reliability. These analyses were followed by an attribution analysis and development of a scoring algorithm. After item reduction, the final version of the questionnaire underwent confirmatory factor analysis, Rasch analysis (only conducted on the unidimensional domain; ie, symptoms), known-groups analysis (only for the symptoms domain because the focus of the instrument was to be able to support a PRO labeling claim for symptoms), and assessments of concurrent and discriminant validity and of internal consistency and test–retest reliability (for symptoms and impacts domains).

In the 12-week study period, patients with RMS were instructed to complete the paper version of the FSIQ-RMS at home (apart from day 1) during 2 intervals: interval 1, from days 1 to 8, where day 1 was used to collect baseline data on the symptoms and impacts items, whereas days 2 to 8 were the first full week of completion of the instrument, and interval 2, from days 80 to 86. Participants were instructed to complete the symptoms domain daily and the impacts domain on days 1, 8, and 86.

Additional PRO questionnaires had to be completed at 3 site visits. Visit 1 occurred on day 1, whereas visits 2 and 3 could occur within a 3-day window after completion of intervals 1 and 2 to accommodate participants' scheduling conflicts. Additional instruments administered included the Medical Outcomes Study RAND-36[29] at visits 1 and 3, and Patient Global Impression of Severity (PGI-S) scales[30] modified for RMS and fatigue and the MFIS[10] at visits 1, 2, and 3. Clinicians completed the Expanded Disability Status Scale (EDSS)[31] at visit 1, and a Clinician Global Impression of Change (CGI-C)[32] at visits 2 and 3.

Control participants without an MS diagnosis completed an adaption of the FSIQ-RMS with all references to RMS removed, at home from days 1 to 7 with a site visit at day 1. Control participants were patients presenting for an acute condition or a checkup and were required to meet the eligibility criteria described in Table 1. Controls were individually matched to RMS patients by sex, age group (in 10-year age cohorts), and race or ethnicity.

Stage 5: real-world study in RMS

Originally developed in US English, the FSIQ-RMS was translated and linguistically validated in 45 language versions (following the Principles of Good Practice for the Translation and Cultural Adaptation Process for PRO Measures[33]), including German. As an

Table 1 – Patient eligibility criteria

Study stage	Inclusion criteria	Exclusion criteria
RMS patients		
All stages	RMS diagnosis as defined by the revised McDonald Diagnostic Criteria for MS Completed written ICF Ambulatory Fluent in English Capable of participating in a 90-minute face-to-face interview	Any condition that may result in energy- or fatigue-related symptoms (other than RMS) Any condition or situation that may interfere significantly with study participation
Stage 1: Concept elicitation interviews Stage 2: Cognitive interviews	Age ≥18 years RRMS diagnosis At least 1 of the following: ≥1 documented relapse in the last 12 months ≥2 documented relapses in the last 24 months EDSS score of 0 to 5.5 in the past 6 months Stable course since EDSS assessment	Current treatment for an autoimmune disorder other than MS
Stage 3: Content confirmation interviews	Age 18 to 65 years PRMS or RSPMS diagnosis EDSS score of 3 to 5.5 in the past 6 months Documented relapse during prior 12 months	Current treatment for an autoimmune disorder other than MS RRMS diagnosis
Stage 4: Psychometric validation	Age 18 to 65 years EDSS score of 0 to 5.5 in the past 1 month or at study entry Documented relapse during past 12 months, excluding the month before baseline visit Medical records available for ≥12 months or since time of RMS diagnosis (if diagnosed during past 12 months)	Current treatment for an autoimmune disorder other than MS History of suicidal ideation Participation in trial with investigational medications for any condition
Stage 6: eDiary migration & usability study	Age 18 to 55 years EDSS score of 0 to 5.5 in the past 3 months Documented relapse(s): ≥1 during past 1 year or ≥2 during past 2 years, excluding the 30 days before screening	Participation in another research study within 30 days before screening Previous participation in any study related to the development of the FSIQ-RMS
Controls		
Stage 4: Psychometric validation	Age 18 to 65 years Fluent in English Completed written ICF	Employed as a shift worker Diagnosis of insomnia, sleep apnea, or any disease that might cause symptoms of fatigue Use of any medication that could cause symptoms of fatigue Any acute disease that is not cured Condition that might interfere significantly with study participation, or predicted to be unable to comply with protocol Participation in trial with investigational medications for any condition

EDSS, Expanded Disability Status Scale; ICF, subject information and consent form including authorization to use and disclose personal health information for research; MS, multiple sclerosis; PRMS, progressive-relapsing MS; RMS, relapsing MS; RRMS, relapsing-remitting MS; RSPMS, relapsing secondary-progressive MS.

addition to a real-world study conducted in Germany and the United States, the FSIQ-RMS was administered to provide data for a supplementary analysis of performance of the symptom items in a real-world setting (n = 154), and to assess differential item functioning in MS patients with depression. As the real-world study was not designed specifically to validate the FSIQ-RMS, it is reported only briefly here, with further details provided in Appendix 1 (see Supplemental Materials found at https://doi.org/10.1016/j.jval.2018.11.007).

Stage 6: eDiary migration and usability study

The final stage was a qualitative migration and usability study in patients with RMS in the United States (n = 10) to demonstrate readability, ease of use, acceptability of an electronic version of the FSIQ-RMS ("eDiary"), and conceptual equivalence between paper and electronic platforms (stage 6).

Before testing, minor changes were made to the FSIQ-RMS to simplify the instructions (these changes were also made to the

paper version), with additional changes to make them appropriate for administration on a handheld electronic device—specifically, a BLU Life Play Android smartphone. The eDiary presented the questionnaire in landscape format with 1 item per screen, providing more room for the horizontal numeric rating scale and space between the responses to facilitate response selection for MS patients who may have problems with fine motor control.

At in-person site visits, participants received training on use of the eDiary, and completed paper and eDiary versions of the FSIQ-RMS and a device-usability questionnaire with 6 questions scored on a verbal rating scale anchored at 1 ("Very easy") and 5 ("Very difficult").

One-on-one cognitive interviews were conducted to probe for participants' interpretation of questions and response options and their perspectives on the usability of the eDiary.

Patient Population

Other than the control participants, all participants in all study stages had RMS diagnoses as defined by the revised McDonald Diagnostic Criteria for MS,[34] with additional subtype diagnostic criteria for the RRMS, PRMS, and RSPMS patients. Recruitment aimed to enroll participants generally representative of the RMS population likely to be enrolled in future clinical trials.

As shown in Table 1, eligibility criteria were highly consistent across study stages. Nevertheless, the upper limit of the age range for eligibility was decreased from 65 years in early study stages to 55 years in later stages to align with the inclusion criteria in 2 phase 3 trials of ponesimod in RMS (OPTIMUM ClinicalTrials.gov identifier NCT02425644 and POINT ClinicalTrials.gov identifier NCT02907177).

Data Analysis

For the qualitative research, participant interviews were audio recorded, transcribed, and anonymized by a third-party transcription agency for qualitative analysis using ATLAS.ti software (ATLAS.ti Scientific Software Development GmbH, Berlin, Germany). Initial coding schemes were developed based on the respective interview guides. Codes were reviewed and modified by the coders as transcripts were analyzed.

In the psychometric validation analyses, item scores were assessed for floor and ceiling effects (ie, the clustering of item scores at the top and bottom of the response-scale range, respectively); a threshold of ≥9% of respondents in the highest or lowest response option was applied. Item discrimination was evaluated using the proportion of participants who endorsed a score for each symptom item within the range of their overall score severity across items (categorized as mild = 0 to 3, moderate = 4 to 6, or severe = 7 to 10), with a threshold of 50%. For inter-item correlation analysis, item pairs with polychoric correlation coefficients of >0.8 were considered to have a high correlation.[35] Because the symptoms domain was considered to measure a unidimensional construct, only symptom items with inter-item correlations greater than 0.90 were reviewed for potential deletion. In exploratory factor analysis, items with factor loadings <0.40 on all factors identified were reviewed for possible deletion.[36] Internal consistency reliability was assessed by calculating the Cronbach's alpha coefficient for each domain and was considered to be met if the Cronbach's alpha coefficient was >0.70.[37] Evaluation of the item hierarchy and ability of the instrument to distinguish between high and low performers used item and person separation indices and item and person reliability indices, with acceptance thresholds of >2.0 for separation indices and >0.8 for reliability indices.[38] Test–retest reliability was assessed using the intraclass correlation coefficient (ICC), calculated using Shrout–Fleiss reliability for a single score, with an ICC of ≥0.70 considered to be acceptable.[39] Unidimensionality was considered to have been demonstrated if >60% of variance was explained by the Rasch factor and <5% was explained by the first residual factor.[40] Sufficient concurrent validity between the new PRO instrument and other instruments was considered if the correlation coefficient was 0.4 to 0.7.[41,42] Descriptive statistical analysis of data for the psychometric analyses was performed using SAS (SAS Institute Inc., Cary, NC, USA). Rasch analyses were performed with Winsteps (Linacre JM, Beaverton, OR, USA. https://www.winsteps.com/winman/references.htm).

Quantitative results are reported as counts and percentages for discrete variables and mean ± standard deviation (SD) for continuous variables.

Results

Participant Characteristics

Characteristics of the participants are reported in Table 2. Across the study stages, participants were predominantly female with mild-to-moderate disease severity, and diverse in age, race/ethnicity, work status (see Appendix 2 in the Supplemental Materials found at https://doi.org/10.1016/j.jval.2018.11.007 for additional employment details), and educational attainment.

Stage 1: Concept Elicitation Interviews in RRMS

Concept elicitation interviews were conducted with a total of 17 participants at 3 US sites in November 2010, and analyses demonstrated that saturation was achieved.

The most commonly reported fatigue-related symptoms were energy and weakness. Functional and emotional impacts of fatigue-related symptoms were reported, including difficulty walking, maintaining relationships, participating in social activities, and performing general activities of daily living, as well as difficulty with thought processes. Elicited symptom and impact concepts are reported in Table 3.

Based on the cognitive debriefing of the FIS, it was determined that this instrument does not sufficiently measure fatigue-related symptoms and their impacts on the RRMS population. The FIS is primarily an impact instrument. Furthermore, most concepts represented in the FIS were reported infrequently or not at all by participants; conversely, several concepts elicited were not present in the FIS.

Based on the results of the concept elicitation interviews and applying the above-mentioned criteria to generate items, a draft version of the new PRO instrument was developed. Instrument drafting also took into account evidence from the literature review and input from clinical experts experienced in treating patients with MS. Reasons for including or not including elicited concepts in the draft instrument are presented in Table 3. Named the "FSIQ-RRMS," the questionnaire comprised 30 items across 2 hypothesized domains: fatigue-related symptoms (n = 16) and impacts of fatigue-related symptoms (n = 14).

Symptom items use an 11-point numeric rating scale assessing the severity of each item, with response options ranging from "Not at all" (0) to "Extremely" (10). This response scale was selected as it is commonly used in clinical settings to assess symptom severity. A recall period of "the past 24 hours" was chosen because symptoms may vary between days.

Impact items use a 5-point verbal rating scale to assess the severity or frequency of each impact, with response options ranging from "No difficulty" to "Extreme difficulty," "Not at all" to "Extremely," or "Never" to "Almost all of the time." A numeric rating scale was not considered appropriate for impacts because items in the impact domain of the questionnaire assess different

Table 2 – Patient demographic and clinical characteristics

Characteristic	Stage 1: Concept elicitation (N = 17)	Stage 2: Cognitive interviews (N = 20)	Stage 3: Content confirmation		Stage 4: Psychometric validation		Stage 6: eDiary migration & usability study (N = 10)
			PRMS (N = 5)	RSPMS (N = 5)	RMS patients (N = 164)	Controls (N = 74)	
RMS subtype, n (%)							
RRMS	17 (100)	20 (100)	–	–	139 (84)	NA	8 (80)
RSPMS	–	–	–	5 (100)	18 (11)	NA	2 (20)
PRMS	–	–	5 (100)	–	8 (5)	NA	0
Age, years, mean ± SD (range)	43.9 ± 13.3 (27 to 75)	47.0 ± 12.0 (25 to 69)	52.6 ± 12.5 (34 to 67)	52.4 ± 10.8 (38 to 63)	45 ± NA (19 to 65)	40 ± NA (18 to 65)	42.1 ± 8.2 (27 to 54)
Female, n (%)	13 (76)	16 (80)	4 (80)	3 (60)	124 (76)	54 (73)	7 (70)
Race/ethnicity, n (%)							
Caucasian/white	13 (76)	17 (85)	2 (40)	4 (80)	133 (81)	56 (76)	5 (50)
African American/ African Caribbean/ Black	2 (12)	3 (15)	3 (60)	1 (20)	NA	NA	4 (40)
Hispanic	0	0	0	0	NA	NA	1 (10)
Asian	1 (6)	0	0	0	NA	NA	1 (10)
American Indian or Alaskan Native	1 (6)	0	0	0	NA	NA	1 (10)
Work status, n (%)*							
Working/studying	8 (47)	8 (40)	2 (40)	0	NA	NA	5 (50)
Not working	9 (53)	12 (60)	3 (60)	5 (100)	NA	NA	5 (50)
Highest level of education, n (%)							
High school diploma or GED	1 (6)	2 (10)	0	1 (20)	47 (29)	13 (18)	2 (20)
Some college or certificate program	4 (24)	8 (40)	2 (40)	2 (40)	50 (30)	26 (35)	4 (40)
College or university degree	12 (71)	10 (50)	3 (60)	2 (40)	66 (40)	34 (46)	4 (40)
Other					1 (1)	1 (1)	0
MS severity†							
Very mild	1 (6)	4 (20)	0	0	–	NA	0
Mild	7 (41)	7 (35)	1 (20)	2 (40)	92 (56)		3 (30)
Moderate	9 (53)	9 (45)	4 (80)	2 (40)	58 (35)		7 (70)
Severe	0	0	0	1 (20)	14 (9)		0
EDSS score, n (%)							
0.0 to 1.0	4 (24)	4 (20)	–	–	23 (14)	NA	1 (10)
1.5 to 2.0	2 (12)	4 (20)	–	–	36 (22)		3 (30)
2.5 to 3.0	4 (24)	4 (20)	0	0	36 (22)		1 (10)
3.5 to 4.0	5 (29)	6 (30)	3 (60)	1 (20)	37 (23)		1 (10)
4.5 to 5.0	1 (6)	0	2 (40)	0	32 (20)		3 (30)
5.5 to 6.0	1 (6)	2 (10)	0	4 (80)			1 (10)

EDSS, Expanded Disability Status Scale; MS, multiple sclerosis; NA, not available; PRMS, progressive-relapsing MS; RMS, relapsing MS; RRMS, relapsing-remitting MS; RSPMS, relapsing secondary-progressive MS.

* Employment details are reported in Appendix 2 in the Supplemental Materials found at https://doi.org/10.1016/j.jval.2018.11.007.

† As assessed by patient for stages 1 to 3, 6: "Very mild," "Mild," "Moderate," or "Severe;" as assessed by PGI-S MS score for stage 4: Mild = 0 to 3, Moderate = 4 to 6, Severe = 7 to 10.

aspects of the patient experience (eg, severity and frequency), and so including verbal prompts for each response choice in the rating scale was considered to be important. A 5-point scale was selected to allow sensitivity in the response choices without adding too much complexity for the patients.[43] A recall period of "past 7 days" was used for all impact items because impacts may not be experienced during each day of a 1-week evaluation period.

Stage 2: Cognitive Interviews in RRMS

Cognitive interviews were conducted with an additional 20 RRMS patients at 3 US sites in February 2011. Overall, the participants reported the draft FSIQ-RRMS instrument to be comprehensive, understandable, and relevant. Nevertheless, participants had a few comments, which were addressed by modifications to the draft questionnaire—namely, removal of 8 symptom items that participants did not interpret as intended (all 6 "at rest" items, which were commonly misinterpreted, and both "exhausted" items, as "exhausted" overlapped with other concepts and was considered by most participants to be a more severe sensation of "tiredness"), revision of the item stem for 6 symptom items to specify that the fatigue symptoms were attributable to RRMS, revision of the item stem for all impact items to specify that the impacts experienced were attributable to fatigue, and rewording of 2 impact items to be

Table 3 – Elicited fatigue-related symptom and impact concepts reported by participants

Concept	Frequency, n (%)	Preliminary decision	Rationale
Symptoms			
Tiredness	15 (88.2)	Exclude	• Instrument will include items specific to physical and mental tiredness. • Concept is not mutually exclusive with "physical" and "mental" tiredness.
Physical tiredness	12 (70.6)	Include	• Concept was frequently reported by participants. • Item divided into 2 separate questions to distinguish between the severity of the symptom "while doing routine daily activities" and "while at rest."
Mental tiredness	6 (35.3)	Include	• Concept was somewhat frequently reported by participants. • Item does not differentiate between "while doing routine daily activities" and "while at rest" as this is not a physical concept.
Fatigue	11 (64.7)	Exclude	• Instrument will include items specific to physical and mental fatigue. • Concept is not mutually exclusive with "physical" and "mental" fatigue.
Physical fatigue	7 (41.2)	Include	• Concept was somewhat frequently reported by participants • Item divided into 2 separate questions to distinguish between the severity of the symptom "while doing routine daily activities" and "while at rest."
Mental fatigue	6 (35.3)	Include	• Concept was somewhat frequently reported by participants • Item does not differentiate between "while doing routine daily activities" and "while at rest" as this is not a physical concept.
Weakness	10 (58.8)	Include	• Concept was somewhat frequently reported by participants. • Concept was solely reported as a physical concept by participants (ie, participants did not report weakness as a mental concept). • Item divided into 2 separate questions to distinguish between the severity of the symptom "while doing routine daily activities" and "while at rest."
Exhaustion	9 (52.9)	Exclude	• Instrument will include items specific to physical and mental exhaustion. • Concept is not mutually exclusive with "physical" and "mental" exhaustion.
Physical exhaustion	6 (35.3)	Include	• Concept was somewhat frequently reported by participants. • Item divided into 2 separate questions to distinguish between the severity of the symptom "while doing routine daily activities" and "while at rest."
Mental exhaustion	6 (35.3)	Include	• Concept was somewhat frequently reported by participants. • Item does not differentiate between "while doing routine daily activities" and "while at rest" as this is not a physical concept.
Low energy	7 (41.2)	Include	• Concept was somewhat frequently reported by participants. • Participants primarily discussed the physical impacts of energy when defining the concept. • Participants typically discussed energy as an amount (depleted, low, or burst of energy).
Worn out/wiped out/run down	6 (35.3)	Include	• Concept was somewhat frequently reported by participants. • Participants used different words to describe this concept; however, "worn out" was most frequently reported (n = 5). • These terms were used to describe both physical and mental concepts; therefore, no further differentiation done.
Sleepiness/Grogginess/ Drowsiness	5 (29.4)	Include	• Concept was somewhat frequently reported by participants. • The term "sleepy" was reported by 3 participants; 1 participant each reported "grogginess" or "drowsiness."
Drained	4 (23.5)	Exclude	• Concept was infrequently reported by participants. • Concept was inconsistently defined by participants.
Heaviness (physical heaviness in limbs)	3 (17.6)	Exclude	• Concept was infrequently reported by participants. • Reports of heaviness were specific to areas of the body that were different for participants (eg, arms, legs, and head were all reported).
Sluggish	2 (11.8)	Exclude	• Concept was infrequently reported by participants.
Decrease in stamina	1 (5.9)	Exclude	• Concept was infrequently reported by participants.
Walking	11 (64.7)	Include	• Frequently reported
Resting/taking breaks	9 (52.9)	Include	• Concept was somewhat frequently reported by participants.
Having to lay down	8 (47.1)	Exclude	• Concept covered by "taking breaks"
Having to sit down	6 (35.3)	Exclude	• Concept covered by "taking breaks"
Unable to do anything/ function physically	6 (35.3)	Exclude	• General concept of being unable to do anything or function physically is too broad to be an appropriate item.
Exercising	5 (29.4)	Exclude	• Concept overlaps with "walking" for 3 out of 5 participants who reference this as an impact. • Concept is not a common experience, as not all patients exercise.
Difficulty standing	3 (17.6)	Exclude	• Concept was infrequently reported by participants.
Lifting	3 (17.6)	Exclude	• Concept was infrequently reported by participants.

continued on next page

Table 3 – *continued*

Concept	Frequency, n (%)	Preliminary decision	Rationale
Need to push oneself	3 (17.6)	Exclude	• Concept was infrequently reported by participants.
Slowed movements	3 (17.6)	Exclude	• Concept was infrequently reported by participants.
Going up/down steps/stairs	2 (11.8)	Exclude	• Concept was infrequently reported by participants.
Carrying	1 (5.9)	Exclude	• Concept was infrequently reported by participants.
Falling	1 (5.9)	Exclude	• Concept was infrequently reported by participants.
Getting in/out of the car	1 (5.9)	Exclude	• Concept was infrequently reported by participants.
Relationships	10 (58.8)	Include	• Concept was somewhat frequently reported by participants. • Participants reported different relationships being impacted, including parent or child (n = 6, 35.3%), with family (n = 3, 17.6%) or social life (n = 3, 17.6%) in general, husband or wife (n = 2), with relation to dating (n = 2, 11.8%), and boy or girlfriend (n = 1, 5.9%).
Social activities	10 (58.8)	Include	• Concept was somewhat frequently reported by participants. • Participants reported a variety of activities, including some that overlap with leisure activities (eg, bowling, dancing, eating out, going to the movies), outdoor activities (eg, going to the zoo with family), and socializing in general (eg, speech).
Planning	5 (29.4)	Exclude	• Concept overlaps with "rearranging plans." The concept as it is presented here is a social concept, but can pertain to general activities of daily living as well.
Difficulty talking with others	3 (17.6)	Exclude	• Concept was infrequently reported by participants. • Concept overlaps with "communicating clearly."
General social impact	2 (11.8)	Exclude	• Concept was infrequently reported by participants. • Participants were nonspecific in their discussion of social impacts with these specific quotes. • Both participants who referenced a social impact with these quotes also reported impacts on specific social activities, and 1 subject each also reported an impact on social relationships and daily planning, and so there is potential overlap with other items measuring "social" concepts (eg, social activities, social relationships, rearranging plans).
Napping	10 (58.8)	Include	• Concept was frequently reported by participants.
Going to bed early	4 (23.5)	Exclude	• Concept was infrequently reported by participants.
Difficulty falling asleep	1 (5.9)	Exclude	• Concept was infrequently reported by participants.
Staying in bed	1 (5.9)	Exclude	• Concept was infrequently reported by participants.
Errands	8 (47.1)	Include	• Concept was somewhat frequently reported by participants. • Participants reported an impact of fatigue-related symptoms on grocery shopping (n = 5, 29.4%), shopping in general (n = 4, 23.5%), and running errands (n = 1, 5.9%). • Item worded to include "running errands" to capture all activities pertinent to the concept, such as "grocery shopping" and "shopping." While "going to the bank or ATM" was not referenced by participants, it is a common experience.
Driving	7 (41.2)	Exclude	• Concept is not a common experience, as not all patients drive a vehicle.
General ADL	7 (41.2)	Exclude	• General concept of ADLs is too broad to be accepted as an appropriate item.
Daily planning	5 (29.4)	Include	• Concept was somewhat frequently reported by participants. • All 5 participants referred to having to plan their day around the severity of their symptoms. Participants mentioned having to "manage" (n = 1), "negotiate" (n = 1), and "rearrange" (n = 1) plans. "Rearrange" was selected as it was a specific impact related to planning, that would also cover the other terms.
Travel	1 (5.9)	Exclude	• Concept was infrequently reported by participants. • Concept is not a common experience, as not all patients travel.
Indoor household activities	8 (47.1)	Include	• Concept was somewhat frequently reported by participants. • Participants reported difficulty performing a variety of activities (eg, washing clothes, ironing, cooking, vacuuming). • Given that there was no consistency over the range of activities reported, no examples were provided.
Outdoor household activities	1 (5.9)	Exclude	• Concept was infrequently reported by participants.

continued on next page

Table 3 – *continued*

Concept	Frequency, n (%)	Preliminary decision	Rationale
Speech	7 (41.2)	Include	• Concept was somewhat frequently reported by participants. • Participants reported several concepts with regard to speech (eg, slurred/slowed speech, blurting out words, talking gibberish) that all speak to difficulty articulating thoughts. • "Communicating clearly" selected as wording to capture all areas of speech that may be affected.
Thought process	7 (41.2)	Include	• Concept was somewhat frequently reported by participants. • While the majority of participants referred to difficulty with their thought "process" or "processing" information (n = 4), "thinking clearly" (n = 1) along with other concepts were reported; "thinking clearly" was selected to simplify the item while making the concept more specific.
Memory	5 (29.4)	Include	• Concept was somewhat frequently reported by participants. • Four participants (80.0%) used the word "forget" when reporting the concept.
Concentration	4 (23.5)	Exclude	• Concept was infrequently reported by participants. • One participant indicated that the difficulty with concentration may have been caused by impaired vision resulting from MS, not fatigue-related symptoms brought on by MS.
Reading	4 (23.5)	Exclude	• Concept was infrequently reported by participants.
Feel less alert	1 (5.9)	Exclude	• Concept was infrequently reported by participants.
Mentally wiped out	1 (5.9)	Exclude	• Concept was infrequently reported by participants. • Nonspecific concept
Frustrated/aggravated/irritated	5 (29.4)	Include	• Concept was somewhat frequently reported by participants. • Feeling "frustrated" was reported most often by participants (n = 3, 17.6%).
Motivation	5 (29.4)	Include	• Concept was somewhat frequently reported by participants.
Anxious/worried/nervous	4 (23.5)	Exclude	• Concept was infrequently reported by participants. • "Anxiety" is a multifaceted term as it contains medical jargon and is also a medical condition.
Feel bad/lousy/miserable	4 (23.5)	Exclude	• Concept was infrequently reported by participants. • General concept of feeling bad is too broad to be accepted as an appropriate item.
Overwhelmed	4 (23.5)	Exclude	• Concept was infrequently reported by participants.
Scared	4 (23.5)	Exclude	• Concept was infrequently reported by participants.
Angry	3 (17.6)	Exclude	• Concept was infrequently reported by participants.
Depressed	2 (11.8)	Exclude	• Concept was infrequently reported by participants. • Concept of "depression" is a multifaceted term, as it is also a medical condition.
Embarrassed	2 (11.8)	Exclude	• Concept was infrequently reported by participants.
Feel sorry for self	2 (11.8)	Exclude	• Concept was infrequently reported by participants.
Disappointed	1 (5.9)	Exclude	• Concept was infrequently reported by participants.
Feel like a burden	1 (5.9)	Exclude	• Concept was infrequently reported by participants.
Lack of interest	1 (5.9)	Exclude	• Concept was infrequently reported by participants.
Overly emotional	1 (5.9)	Exclude	• Concept was infrequently reported by participants.
Sad	1 (5.9)	Exclude	• Concept was infrequently reported by participants.
Bathing/washing	5 (29.4)	Include	• Concept was somewhat frequently reported by participants.
Productivity	5 (29.4)	Exclude	• Concept is not a common experience, as not all patients with MS are working.
Schedule	5 (29.4)	Exclude	• Concept is not a common experience, as not all patients with MS are working or going to school. • Overlaps with "rearranging plans"
General work/school impact	4 (23.5)	Exclude	• Concept was infrequently reported by participants. • Concept is not a common experience, as not all patients with MS are working. • General concept of work or school is too broad to be accepted as an appropriate item.
Not working	4 (23.5)	Exclude	• Concept was infrequently reported by participants. • Concept is not a common experience, as only some patients with MS are working.

continued on next page

Table 3 – *continued*

Concept	Frequency, n (%)	Preliminary decision	Rationale
Attendance	2 (11.8)	Exclude	• Concept was infrequently reported by participants. • Concept is not a common experience, as not all patients with MS are working. • Overlaps with "rearranging plans"
Relationships with coworkers	2 (11.8)	Exclude	• Concept was infrequently reported by participants. • Concept is not a common experience, as not all patients with MS are working.
School	2 (11.8)	Exclude	• Concept was infrequently reported by participants. • Concept is not a common experience, as not all patients with MS are in school.
Job security	1 (5.9)	Exclude	• Concept was infrequently reported by participants. • Concept is not a common experience, as not all patients with MS are working. • Broad concept that would be a secondary impact to other affected areas of work (ie, fatigue-related symptoms of RRMS would not directly impact job security; impact of fatigue-related symptoms of RRMS on productivity, schedule, relationships with coworkers, etc, would impact job security).
Active leisure activities	4 (23.5)	Exclude	• Concept was infrequently reported by participants.
Sedentary leisure activities	3 (17.6)	Exclude	• Concept was infrequently reported by participants.
Sex life	2 (11.8)	Exclude	• Concept was infrequently reported by participants. • Concept is not a common experience, as not all patients are sexually active.
Restless legs	1 (5.9)	Exclude	• Concept was infrequently reported by participants.
Sitting	1 (5.9)	Exclude	• Concept was infrequently reported by participants.
Financial impact	3 (17.6)	Exclude	• Concept was infrequently reported by participants.

ADL, activities of daily living; MS, multiple sclerosis; RRMS, relapsing-remitting MS.
Frequently reported, ≥11 participants; somewhat frequently reported, 5 to 10 participants; infrequently reported, ≤4 participants.

more specific. No revisions were required to the recall period or response options. After revision, the PRO instrument comprised 22 items: 8 fatigue-related symptoms and 14 impacts.

Subsequently, a Rasch analysis was conducted on the participant responses collected during the stage 2 cognitive interviews. Based on this analysis, a previously deleted symptom item was added back to the instrument to ensure that symptoms relevant for the severe spectrum of the disease were adequately covered, resulting in 23 items in total.

Stage 3: Content Confirmation Interviews in PRMS, RSPMS

A total of 10 patients with PRMS (n = 5) or RSPMS (n = 5) were interviewed at 2 US sites in May 2012. In the concept elicitation portion, new concepts were reported by only 1 or 2 participants, overlapped with already-included concepts, or were too nonspecific for inclusion in the questionnaire (eg, unable to do anything/function physically, being in a "bad mood"). Consequently, no items were added (nor were any other modifications made) to the draft 23-item questionnaire, which was renamed the "FSIQ-RMS v1."

Stage 4: Psychometric Validation in Observational Study

Results of the psychometric analyses are summarized in Table 4. In total, data were analyzed for 164 patients with RMS recruited from 15 US sites and 74 controls recruited from 3 US sites, between September 2012 and October 2013.

Instrument completion

Only 3 participants had 1 or more missing responses on symptom items by day 8, and no symptom item had data missing from more than 1 participant, indicating that missing data are a random effect. The FSIQ-RMS impact items also had low levels of missing data.

Item performance analysis

Participants used the full range of the symptoms scale (0 to 10) and the impacts scale (0 to 4), with similar distributions for all items. Substantial ceiling effects were observed for most symptom and impact items, reflecting the mild-to-moderate patient population with few ambulatory limitations. Symptom items without ceiling effects on day 1 and/or day 8 were "Physically tired," "Physically fatigued," and "Energy doing routine daily activities." The only impact item without ceiling effects on visit 2 (day 9 to 11) was, "How often did you have to take a break?"

High pairwise inter-item correlations (≥0.94) on days 1 and 8, potentially indicating redundancy between content-related item pairs, were observed for the symptom items "Mentally tired" and "Mentally fatigued" and for "Physically tired" and "Physically fatigued." No impact item pairs had correlations >0.8 on both days 1 and 8.

The proportion of participants who endorsed a score for each symptom item within the range of their overall score severity across items was consistently above the 50% threshold, indicating that the instrument is able to discriminate between mild, moderate, and severe fatigue.

Factor analysis

In exploratory factor analysis of the 9-item symptoms domain, 1 factor explained 93.4% of the variance (Eigenvalue = 47.74; $P < .0001$), supporting the hypothesis that the symptom items reflect a single underlying construct. Exploratory factor analysis of the 14-item impacts domain resulted in 3 factors explaining 9.68%, 8.37%, and 4.94% of the variance, respectively

Table 4 – Measurement properties of the new PRO in the psychometric study.

Parameter	Threshold for acceptability	FSIQ-RMS symptoms domain (9-item), observational study, day 1 (N = 164)	FSIQ-RMS symptoms domain (7-item), observational study, day 1 (N = 164)	FSIQ-RMS impacts domain (14-item), observational study, day 1 (N = 164)	FSIQ-RMS impacts domain (13-item), observational study, day 1 (N = 164)
Rasch analysis					
Person reliability	≥0.8	0.93	0.91	0.91	–
Person separation	≥2.0	3.57	3.19	3.12	–
Item reliability	≥0.8	0.90	0.92	0.98	–
Item separation	≥2.0	3.07	3.49	6.32	–
Dimensionality, %	≥60	76.3	75.0	59.4	–
Item misfit	0.5 to 1.5	No item misfit by day 8. Slight misfit for the item "worn out at rest" on day 1	No item misfit	No item misfit	–
Residual correlations	≤0.40	Mentally tired and mentally fatigued (0.66) Physically tired and physically fatigued (0.52)	No residual correlations	No residual correlations	–
Measure range		−0.21 to 0.38	−0.20 to 0.37	−1.26 to 1.78	–
Rating scale		No disordered steps between adjacent categories	No disordered steps between adjacent categories	No disordered steps between adjacent categories	–
Classical statistics					
Item response and dimensionality analysis					
Complete data		98.2%	Evaluated in Stage 1 to make item deletion decisions. Not evaluated on the final measure as the results would be the same.	–	
Ceiling range	≤9%	3.7% to 16.5%		–	
Floor range	≤9%	1.8% to 9.8%		–	
Inter-item correlations	≥0.90	Mentally tired and mentally fatigued (0.95) Physically tired and physically fatigued (0.95)		None	
Item discrimination	≥50%	52% to 60%		–	
Factor analysis, explained variance	≥40%	93.4% (Eigenvalue = 47.74)		9.68%, 8.37%, and 4.94% (Eigenvalue = 19.92, 1.88, and 1.18, respectively)	
Psychometric testing of final questionnaire (reliability)					
Internal consistency reliability	≥0.7		0.949		Physical: 0.869 Cognitive/ Emotional: 0.879 Coping: 0.872
Test–retest reliability	≥0.7		0.928		Physical: 0.946 Cognitive/ Emotional: 0.917 Coping: 0.945
Known-groups validity	$P < .05$ for difference in scores between known groups		Significant difference in mean scores: Patients: 44.63 Healthy subjects: 17.67		

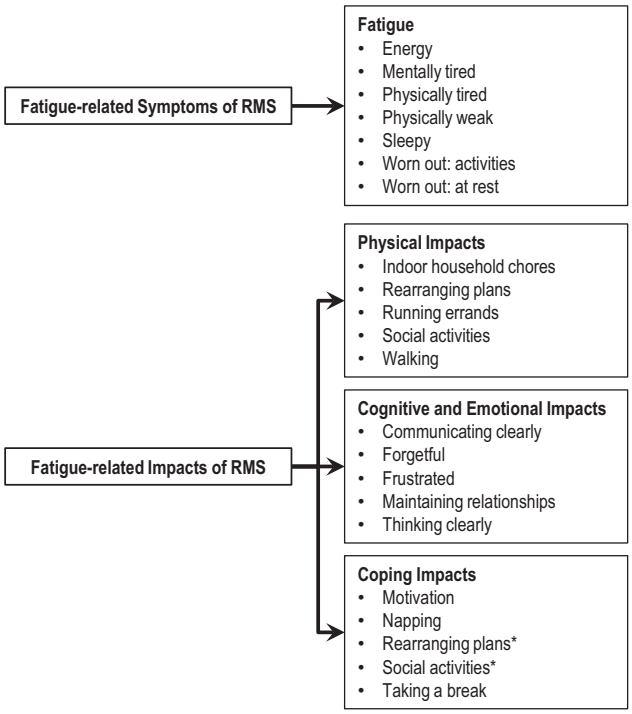

Fig. 2 – Conceptual framework. RMS, relapsing multiple sclerosis. *Items also present in physical impacts subdomain.

(Eigenvalue = 19.92, 1.88, and 1.18, respectively; $P < .0001$), suggesting that the impact items can be categorized into 3 conceptual subdomains.

Rasch analysis of the 9-item symptoms domain revealed that a unidimensional factor explaining most of the variance was "fatigue." Rasch analysis of the 14-item impacts domain provided evidence of multidimensionality, with the Rasch dimension explaining 59.4% and 58.4% of the variance on days 1 and 8, respectively, thus not reaching the threshold for demonstrating unidimensionality.

Item reduction

Based on the observed redundancy between symptom items measuring "fatigue" and "tired," the items "Physically fatigued" and "Mentally fatigued" were deleted.

The impacts item related to showering was deleted because it had a substantial ceiling effect, with 57.9% of participants responding that they had no difficulty showering. There was also categorical disordering within the scale, indicating that patients do not use the response options as intended. In addition, this item may not be generally applicable in a multinational study because showers may not be available in all countries.

As a result of these modifications, the final FSIQ-RMS v2 incorporated 1 hypothesized fatigue-related symptoms domain comprising 7 items, and 3 hypothesized impacts subdomains—physical, cognitive/emotional, and coping (which is combined with 2 physical impacts subdomain items)—comprising 13 items (Fig. 2).

Scoring algorithm

A scoring algorithm was developed to standardize scores on both the symptoms domain (daily and weekly) and impacts subdomains (weekly) to a 0 to 100 scale, with higher scores indicating more severe symptoms and impacts. There is no single summary score across the FSIQ-RMS instrument, but rather 1 symptoms score and 3 impacts subdomain scores.

For the average weekly symptoms domain score to be computed, data from at least 4 daily diaries with at least 4 items (ie, ≥50% nonmissing item responses) completed on each day are needed. If fewer than 4 items are reported on a day, the day is considered missing. If fewer than 4 days are available within the 7-day period, the observation for the weekly average score is considered missing.

For the impacts domain, the 2 items present in both the physical and coping subdomains ("taking part in social activities" and "rearranging plans") are multidimensional, and represent coping strategies related to the physical impact of the disease. For each impacts subdomain, at least 3 items (ie, ≥50% nonmissing item responses) are needed to calculate the respective weekly subdomain score. If fewer than 3 items are reported, the observation for that week is considered missing.

Psychometric testing of final FSIQ-RMS

In confirmatory factor and Rasch analyses, the final FSIQ-RMS 7-item symptoms domain demonstrated unidimensionality, with the model explaining 75.0% of the variance on day 1 and 83.2% on day 8. The symptoms domain was shown to contain independent, reliable items, with item reliability indices ranging from 0.92 to 0.95 across items, exceeding the threshold for demonstrating score repeatability. Item scores were able to discriminate patients along the severity continuum, with item separation indices ranging from 3.49 to 4.35, exceeding the threshold for demonstrating sensitivity of the instrument.

In analyses of concurrent validity, a higher FSIQ-RMS symptoms score was—as expected—associated with lower perceived health as measured by the RAND-36 Energy/fatigue scale on day 1, and higher fatigue symptoms as measured by the MFIS Physical subscale on days 1 and 2 to 8 (Table 4). All 3 impacts subdomains were significantly correlated with the RAND-36 Role functioning/physical scale on day 1, and the MFIS Physical and Psychosocial subscales on days 1 and 8.

Known-groups validity was demonstrated by symptoms domain scores differing significantly between RMS patients and controls (Table 4).

All symptoms and impacts subdomains of the FSIQ-RMS exceeded the threshold for internal consistency, with Cronbach's alpha on day 1 ranging from 0.869 to 0.949. All symptoms and impacts subdomains also exceeded the threshold for demonstrating test–retest reliability, with ICCs ranging from 0.917 to 0.946.

Attribution analysis

The differential item functioning analysis, which was performed on the symptoms domain, found greater impairment on 3 items in patients with RMS compared with controls, confirming significantly worse fatigue in RMS patients compared with non-RMS individuals.

In a regression model controlling for age, cognition, and level of depression, there was a significant 14-point difference in symptoms score (scale range 0 to 100) between RMS patients and controls ($P < .0001$), demonstrating that the fatigue-related symptoms were attributable to RMS.

Stage 5: Real-World Study

Of 1597 MS patients for whom data were gathered in 2012, 154 completed the FSIQ-RMS symptoms measure and met the eligibility criteria for analysis. Results confirmed that the FSIQ-RMS symptoms domain items met the unidimensionality and local

independence assumptions of the Rasch model, and supported the item reductions described above (see Appendix 1 in the Supplementary Materials found at https://doi.org/10.1016/j.jval.2018.11.007 for further details).

Stage 6: eDiary Migration and Usability Study

Data were analyzed for a total of 10 participants from 2 US sites who were interviewed in January 2015. On the device-usability questionnaire, all participants reported that it was easy or very easy to learn, use, navigate, read, and see the response choices of the eDiary.

In interviews, 7 participants (70%) spontaneously reported preferring the eDiary to the paper format, based on better ease of use (n = 4), less information per page (n = 2), being quicker and simpler to use (n = 3), not having to write (n = 1), and providing better flow (n = 1). Participants generally interpreted the questionnaire the same way on the 2 platforms.

In the cognitive interviews, participants confirmed that the symptoms and impacts listed on the FSIQ-RMS were relevant to their experience with RMS. Participants generally reported understanding the meaning of the questions, instructions, and response options, though 2 participants mentioned confusion about the meaning of "at rest" in 1 question. After stage 6, examples were added to this question: "eg, reading a book, watching TV."

Discussion

Because fatigue is the most frequent symptom of MS and has a severe impact on patients' health-related quality of life, accurate measurement of fatigue is important in devising the best treatment strategies for each patient.[44] The final FSIQ-RMS is a valid and reliable PRO instrument that has demonstrated content and measurement validity for fatigue-related symptom and impact items. The multistage development and psychometric validation program established that the new questionnaire measures important aspects of fatigue symptoms attributable to relapsing forms of MS and the impacts resulting from these symptoms.

Development and validation of the FSIQ-RMS were designed from the outset to be rigorous with respect to current requirements for PRO instrument development. Recently, other PRO instruments for assessing fatigue in patients with MS have been developed following different approaches; for example, the Neurological Fatigue Index (NFI-MS) comprises a 10-item summary scale, and scales measuring the physical and cognitive aspects of fatigue.[19] In contrast to the FSIQ-RMS, the NFI-MS is based on a fatigue-symptom framework.[45]

Although the FSIQ-RMS has several strengths, including rigorous development following the FDA PRO guidance,[12] incorporation of the input of RMS patients, and psychometric validation performed in a stand-alone study that included controls to confirm attribution of symptoms to RMS, the instrument development also has limitations. Because of the mild-to-moderate disease of most participants, the validity of the FSIQ-RMS in patients with more severe disease (ie, those with EDSS >5.5) would have to be established in future research. As the instrument has not been assessed in MS subtypes other than RRMS, PRMS, and RSPMS, it cannot be assumed to be applicable to patients with different disease subtypes.

In conclusion, qualitative research and psychometric validation analyses indicate that the FSIQ-RMS is fit for purpose for measuring fatigue symptoms and impacts as a clinical study endpoint in RMS trials. Results of the migration and usability study confirmed the conceptual equivalence of the FSIQ-RMS eDiary to the paper version and its appropriateness for use with RMS patients. Responsiveness and meaningful change, as well as variability by language version and demographic subgroups (eg, sex, country), of the final FSIQ-RMS eDiary will be analyzed in 2 international phase 3 trials of ponesimod: OPTIMUM (ponesimod vs teriflunomide in patients with active RMS; NCT02425644) and POINT (add-on ponesimod to dimethyl-fumarate in patients with active RMS despite treatment with dimethyl-fumarate; NCT02907177). First results, coming from the OPTIMUM trial, are anticipated in 2019.

Role of the Funding Source

This research was funded by Actelion Pharmaceuticals Ltd, Allschwil, Switzerland. For the qualitative study, patient recruitment was conducted by Global Market Research Group and MedQuest, and participant interviews and data analysis were conducted by Adelphi Values; the psychometric study was conducted by Adelphi Values; the real-world study was conducted by Adelphi RealWorld; data analysis for the psychometric and real-world studies was conducted by Clinical Outcomes Solutions; the ePRO migration study was conducted by Evidera; the handheld ePRO device was programmed by CRF Health, Inc.—all funded by Actelion. Actelion was involved in study design; in the analysis and interpretation of data; in the writing of the article; and in the decision to submit the article for publication.

Acknowledgments

Andrew Yaworsky from Adelphi Values contributed to the qualitative study; Adelphi Values received funding from Actelion for this research. William Gehringer from Clinical Outcomes Solutions and Benjamin Banderas from Adelphi Values participated in the psychometric validation study; Clinical Outcomes Solutions received funding from Actelion for this research. Shannon Shaffer from Evidera contributed to the ePRO migration study; Evidera received funding from Actelion for this research. Medical writing and editorial support were provided by W. Mark Roberts, PhD, Montréal, Québec, Canada, funded by Actelion. Portions of this work were previously presented at the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) 19th Annual International Meeting, Montréal, Canada, June 3, 2014; the ISPOR 20th Annual International Meeting, May 18, 2015, Philadelphia, PA, USA; and the ISPOR 18th Annual European Congress, November 11, 2015, Milan, Italy.

Conflict of Interest

S. Hudgens is a salaried employee of Clinical Outcomes Solutions; Clinical Outcomes Solutions has received payments from Actelion Pharmaceuticals Ltd (Switzerland) for services related to the conduct of this research. R. Schüler is an employee of Actelion and previously held stock options/bond holdings in Actelion. E. Hunsche was an employee of Actelion at the time this research was conducted (current affiliation: Myovant Sciences GmbH) and previously held stock options/bond holdings in Actelion. J. Stokes is a salaried employee of Adelphi Values; Adelphi Values has received payments from Actelion for services related to the conduct of this research. S. Eremenco was a salaried employee of Evidera at the time this research was conducted (current affiliation: C-Path); Evidera has received payments from Actelion for services related to

the conduct of this research. T. P. Leist reports no relevant conflicts of interest.

Author Contributions

S. Hudgens, R. Schüler, J. Stokes, S. Eremenco, and E. Hunsche contributed to study design. All authors participated in study conception, data analysis, data interpretation, article drafting and critical revision, and approved the final article. All authors were involved in the decision to submit the article for publication. All authors had full access to all of the data in this study and take complete responsibility for the integrity of the data and the accuracy of the data analysis.

Supplemental Materials

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.jval.2018.11.007.

REFERENCES

1. Patrick E, Christodoulou C, Krupp LB, New York State MS Consortium. Longitudinal correlates of fatigue in multiple sclerosis. *Mult Scler*. 2009;15:258–261.
2. Fisk JD, Pontefract A, Ritvo PG, et al. The impact of fatigue on patients with multiple sclerosis. *Can J Neurol Sci*. 1994;21:9–14.
3. Williams AE, Vietri JT, Isherwood G, Flor A. Symptoms and association with health outcomes in relapsing-remitting multiple sclerosis: results of a US patient survey. *Mult Scler Int*. 2014;2014:203183.
4. Dendrou CA, Fugger L, Friese MA. Immunopathology of multiple sclerosis. *Nat Rev Immunol*. 2015;15:545–558.
5. Smith MM, Arnett PA. Factors related to employment status changes in individuals with multiple sclerosis. *Mult Scler*. 2005;11:602–609.
6. Glanz BI, Degano IR, Rintell DJ, et al. Work productivity in relapsing multiple sclerosis: associations with disability, depression, fatigue, anxiety, cognition, and health-related quality of life. *Value Health*. 2012;15:1029–1035.
7. Flensner G, Landtblom AM, Soderhamn O, Ek AC. Work capacity and health-related quality of life among individuals with multiple sclerosis reduced by fatigue: a cross-sectional study. *BMC Public Health*. 2013;13:224.
8. Janardhan V, Bakshi R. Quality of life in patients with multiple sclerosis: the impact of fatigue and depression. *J Neurol Sci*. 2002;205:51–58.
9. Lerdal A, Celius EG, Moum T. Fatigue and its association with sociodemographic variables among multiple sclerosis patients. *Mult Scler*. 2003;9:509–514.
10. MS Council for Clinical Practice Guidelines. *Fatigue and Multiple Sclerosis: Evidence-Based Management Strategies for Fatigue in Multiple Sclerosis*. Washington, DC: Paralyzed Veterans of America; 1998.
11. Rae-Grant A, Bennett A, Sanders AE, et al. Quality improvement in neurology: multiple sclerosis quality measures: executive summary. *Neurology*. 2015;85:1904–1908.
12. Food and Drug Administration. Guidance for Industry. Patient-reported outcome measures: use in medical product development to support labeling claims. http://www.fda.gov/downloads/Drugs/Guidances/UCM193282.pdf. Accessed June 16, 2014.
13. Food and Drug Administration. Roadmap to patient-focused outcome measurement in clinical trials. http://www.fda.gov/Drugs/DevelopmentApprovalProcess/DrugDevelopmentToolsQualificationProgram/ucm370177.htm. Accessed June 18, 2014.
14. Walton MK, Powers 3rd JH, Hobart J, et al. Clinical outcome assessments: conceptual foundation—report of the ISPOR Clinical Outcomes Assessment—Emerging Good Practices for Outcomes Research Task Force. *Value Health*. 2015;18:741–752.
15. Krupp LB, LaRocca NG, Muir-Nash J, Steinberg AD. The fatigue severity scale. Application to patients with multiple sclerosis and systemic lupus erythematosus. *Arch Neurol*. 1989;46:1121–1123.
16. Fisk JD, Ritvo PG, Ross L, et al. Measuring the functional impact of fatigue: initial validation of the fatigue impact scale. *Clin Infect Dis*. 1994;18(Suppl 1):S79–S83.
17. Meads DM, Doward LC, McKenna SP, et al. The development and validation of the Unidimensional Fatigue Impact Scale (U-FIS). *Mult Scler*. 2009;15:1228–1238.
18. Penner IK, Raselli C, Stocklin M, et al. The Fatigue Scale for Motor and Cognitive Functions (FSMC): validation of a new instrument to assess multiple sclerosis-related fatigue. *Mult Scler*. 2009;15:1509–1517.
19. Mills RJ, Young CA, Pallant JF, Tennant A. Development of a patient reported outcome scale for fatigue in multiple sclerosis: The Neurological Fatigue Index (NFI-MS). *Health Qual Life Outcomes*. 2010;8:22.
20. Barlow S, Brimelow R, Burke T, et al. Clinical evaluation of fatigue in MS using the Brief Fatigue and Associated Symptoms Scale. *Br J Neurosci Nurs*. 2016;12:216–227.
21. Mills R, Young C, Nicholas R, et al. Rasch analysis of the Fatigue Severity Scale in multiple sclerosis. *Mult Scler*. 2009;15:81–87.
22. Mills RJ, Young CA, Pallant JF, Tennant A. Rasch analysis of the Modified Fatigue Impact Scale (MFIS) in multiple sclerosis. *J Neurol Neurosurg Psychiatry*. 2010;81:1049–1051.
23. Lublin FD, Reingold SC, Cohen JA, et al. Defining the clinical course of multiple sclerosis: the 2013 revisions. *Neurology*. 2014;83:278–286.
24. Compston A, Coles A. Multiple sclerosis. *Lancet*. 2008;372:1502–1517.
25. Tremlett H, Zhao Y. Primary and secondary progressive MS have a similar age at onset of progression—NO. *Mult Scler*. 2017;23:640–642.
26. Doward LC, Meads DM, Fisk J, et al. International development of the Unidimensional Fatigue Impact Scale (U-FIS). *Value Health*. 2010;13:463–468.
27. Newland P, Starkweather A, Sorenson M. Central fatigue in multiple sclerosis: a review of the literature. *J Spinal Cord Med*. 2016;39:386–399.
28. Lamoureux RE, Shields A, Stokes J, et al. How many subjects are enough for symptom-focused concept elicitation studies? A retrospective analysis of saturation across twenty-six studies. *Value Health*. 2015;18:A33.
29. Hays RD, Sherbourne CD, Mazel RM. The RAND 36-Item Health Survey 1.0. *Health Econ*. 1993;2:217–227.
30. Yalcin I, Bump RC. Validation of two global impression questionnaires for incontinence. *Am J Obstet Gynecol*. 2003;189:98–101.
31. Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). *Neurology*. 1983;33:1444–1452.
32. Guy W. *ECDEU Assessment Manual for Psychopharmacology*. Washington, DC: Department of Health, Education, and Welfare; 1976.
33. Wild D, Grove A, Martin M, et al. Principles of Good Practice for the Translation and Cultural Adaptation Process for Patient-Reported Outcomes (PRO) Measures: report of the ISPOR Task Force for Translation and Cultural Adaptation. *Value Health*. 2005;8:94–104.
34. Polman CH, Reingold SC, Edan G, et al. Diagnostic criteria for multiple sclerosis: 2005 revisions to the "McDonald Criteria". *Ann Neurol*. 2005;58:840–846.
35. Olsson U, Drasgow F, Dorans NJ. The polyserial correlation coefficient. *Psychometrika*. 1982;47:337–347.
36. Ferguson E, Cox T. Exploratory factor analysis: a users' guide. *Int J Select Assess*. 1993;1:84–94.
37. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika*. 1951;16:297–334.
38. Van Zile-Tamsen C. Using Rasch analysis to inform rating scale development. *Res High Educ*. 2017;58:922–933.
39. Kaiser HF. The application of electronic computers to factor analysis. *Educ Psychol Meas*. 1960;20:141–151.
40. Linacre JM. Data variance: explained, modeled and empirical. *Rasch Meas Trans*. 2003;17:942–943.
41. Litwin MS. *How to Measure Survey Reliability and Validity*. Thousand Oaks, CA: Sage; 1995.
42. Nunnally JC. The assessment of reliability. In: Bernstein I, ed. *Psychometric Theory*. New York: McGraw Hill; 1994.
43. Nagata C, Ido M, Shimizu H, et al. Choice of response scale for health measurement: comparison of 4, 5, and 7-point scales and visual analog scale. *J Epidemiol*. 1996;6:192–197.
44. Induruwa I, Constantinescu CS, Gran B. Fatigue in multiple sclerosis—a brief review. *J Neurol Sci*. 2012;323:9–15.
45. Mills RJ, Young CA. A medical definition of fatigue in multiple sclerosis. *QJM*. 2008;101:49–60.