Patient-reported outcome measures in MS: Do development processes and patient involvement support valid quantification of clinically important variables?

Bharadia, Trishna; Vandercappellen, Jo; Chitnis, Tanuja; Eelen, Piet; Bauer, Birgit; Brichetto, Giampaolo; Lloyd, Andrew; Schmidt, Hollie; King, Miriam; Fitzgerald, Jennifer; Hach, Thomas; Hobart, Jeremy

Published in:
Multiple Sclerosis Journal - Experimental, Translational and Clinical

DOI:
10.1177/20552173221105642

Publication date:
2022

MSJ Experimental Translational Clinical

Original Research Article

Patient-reported outcome measures in MS: Do development processes and patient involvement support valid quantification of clinically important variables?

Multiple Sclerosis Journal—Experimental, Translational and Clinical

April–June 2022, 1–18

DOI: 10.1177/20552173221105642

Trishna Bharadia*, Jo Vandercappellen*, Tanuja Chitnis, Piet Eelen, Birgit Bauer, Giampaolo Brichetto, Andrew Lloyd, Hollie Schmidt, Miriam King, Jennifer Fitzgerald, Thomas Hach* and Jeremy Hobart*

Abstract

Background: Patient-reported outcomes (PROs) are widely measured in multiple sclerosis (MS) studies. However, the quality of instrument development processes varies, raising concerns about the meaningfulness of associated data.

Objectives: To review the development of selected PROs commonly used in MS studies, including definitions of the concepts measured, use of conceptual frameworks, and degree of input from people living with MS (PlwMS). To gain insights and recommendations from PlwMS on their experience with these PROs.

Methods: We assessed 6 PROs (FSIQ-RMS, modified-FIS, MSQoL-54, Leeds 8-item MSQoL, MSIS-29 and EQ-5D) for alignment with regulatory and scientific requirements on PRO structure/development. PlwMS evaluated the degree to which the PROs reflect disease aspects they perceive important.

Results: Definitions, clarifications and conceptualisations of the measurement variables were often lacking. PlwMS were variably involved in PRO development. Ethnic diversity was rarely documented. PlwMS identified individualisation, ease of understanding, time burden, and mode of administration as factors affecting PRO usability.

Conclusions: To date, the PRO development process has consistently lacked clear definitions of concepts of interest, use of conceptual frameworks and patient involvement, thereby compromising the validity of data they generate. PRO instrument development must be conducted more robustly to maximise the value of pivotal clinical trials.

Keywords: Multiple sclerosis, patient-reported outcomes, fatigue, symptoms, impact, insights

Date received: 14 January 2022; accepted: 21 May 2022

Introduction

People living with Multiple sclerosis (PlwMS) have a range of symptoms impacting their life quality.[1,2] Some can be measured objectively, others require assessment and quantification of patient's perceptions.[3] Patient-reported outcome (PRO) measures seek to provide these data.[4]

PROs play key roles in MS studies.[5] They quantify MS impacts, monitor these longitudinally, determine therapeutic and cost effectiveness, and interpret the clinical meaningfulness of changes in objective measures. Decisions based on PRO interpretations influence the lives of PlwMS, health care utilisation, and public expenditure. It is hard to construct an argument to support compromising PRO quality.

Guidance for PRO development and selection for clinical trials is evolving. Recent guidance highlight the importance of conceptual frameworks, patient involvement,

Correspondence to: **Jeremy Hobart**, Peninsula Schools of Medicine and Dentistry, University of Plymouth, Plymouth, UK. **jeremy.hobart@plymouth.ac.uk**

*equally contributing authors

Trishna Bharadia, Patient Author, Marlow, UK

Jo Vandercappellen, Novartis Pharma AG, Basel,

Switzerland

Tanuja Chitnis, Department of Neurology, Brigham and Women's Hospital, Boston, MA, USA

Piet Eelen, National Multiple Sclerosis Center of Melsbroek, Flanders, Belgium

Birgit Bauer, Manufaktur für Antworten UG, Abensberg, Germany

Giampaolo Brichetto, Associazione Italiana Sclerosi Multipla Rehabilitation Center, Genoa, Italy

Andrew Lloyd, Acaster Lloyd Consulting Ltd, London, UK

Hollie Schmidt, Accelerated Cure Project for Multiple Sclerosis, Waltham, MA, USA

Miriam King, Novartis Pharma AG, Basel, Switzerland

Jennifer Fitzgerald, Novartis Pharma AG, Basel, Switzerland

Thomas Hach, Novartis Pharma AG, Basel, Switzerland

Jeremy Hobart, Peninsula Schools of Medicine and Dentistry, University of Plymouth, Plymouth, UK

and advanced psychometric methods.[4,6,7] Conceptual frameworks are the hypothesised relationships between a measurement variable (clinical concept), its components, their sub-components, proposed scale items, and scores generated. They are not just important, they underpin PRO validity.[8] Our experience is that these recommendations and advances are not yet fully appreciated by MS clinical trialists.

The PROs that matter to PlwMS initiative (PROMPT-MS) aims to highlight the need for a new generation of robustly designed PROs. Specifically, to emphasise the need for clear and specific PRO concept definitions, promote PRO selection strategies, highlight alignment with regulatory and best available science guidance and, most importantly, what really matters to PlwMS.

This study had two aims. First, to review the development of selected commonly used PROs, with respect to PlwMS's involvement, concept definitions, and conceptual frameworks used. Second, to gain insights from PlwMS in relation to their experiences of completing PROs, how effectively they perceive these instruments capture factors relevant to PlwMS, and the strengths and weaknesses of selected PRO instruments.

Methods

Evaluation of PRO development

A published literature review[9] combined with guidance from the PROMPT-MS Steering Committee (**Supplementary Table 1**) identified 6 PROs for evaluation that were considered to be representative of available PROs. Two measured fatigue: the Fatigue Symptoms and Impacts Questionnaire – Relapsing MS (FSIQ-RMS)[10] and 21-item modified Fatigue Impact Scale (mFIS).[11,12] Three measured 'quality of life': 54-item MS QoL scale (MSQoL-54);[13] 8-item Leeds MS QoL measure (LMSQoL),[14] and EuroQol EQ-5D.[15] The final, 29-item MS Impact Scale (MSIS-29),[16] measured the physical and psychological impact of MS. The selected instruments are not intended to constitute an exhaustive list of PROs. We chose a small set of relevant instruments designed to assess a range of important patient-centric symptoms that are challenging domains to define accurately. The selections of the mFIS and FSIQ-RMS were to contrast older and newer PROs (i.e. instruments developed before and after the publication of regulatory guidelines).

We assessed the PRO development process from the original PRO development papers, extracting information on instrument purpose, concept and domain definitions, conceptual frameworks, patient input, and instrument items and response formats. We developed a profile for each instrument (Figure 1).

PROs: qualitative insights from PlwMS

PlwMS ($n = 25$) were interviewed to gain insights on their prior experience of using PROs, perceived appropriateness of PRO item response options, opinions of suitable recall periods, the degree to which scores accurately represent concepts being measured, and views on alternative approaches to assessing the impact of MS. These insights complemented the profiling exercise described above. The PlwMS who were interviewed were all actively involved and vocal members of the MS community, and many shared their experiences on social media. Table 1 shows their characteristics. Most were females (72%) and had relapsing-remitting MS (92%). A range of nationalities, ethnicities, ages, disease durations, and relationship statuses were represented. Both interviewer and interviewees had MS, to help interviewees feel at ease, to ensure that interviewers were subject matter experts, and to embed the patient perspective throughout the research process. The interviewer was trained under the European Patients' Academy on Therapeutic Innovation (EUPATI) initiative (**Supplementary Materials**)[17] and was accompanied by a qualified moderator (market research expert).

Interviewees also completed a post-interview survey on the strengths and weaknesses of the six PROs. The purpose of the post-interview survey was to capture additional information not gathered during the interviews. We recognised the value of a period of reflection following the interviews. The survey enabled participants to have a second opportunity to comment on specific issues after a period of reflection. Specifically, the post-interview survey related to patients' perceptions of PRO strengths and weaknesses, and whether they felt the tools effectively assessed the impact of MS on aspects of their life quality, whether they covered aspects that are relevant to people living with MS, whether they would create a significant time burden, and whether they would be manageable to complete. The interview schedule and questionnaire are available in the **Supplementary Material**. All interviewees provided written informed consent. Data analyses were conducted by a registered psychologist and psychotherapist and were compliant with the EphMRA market research code of conduct.[18]

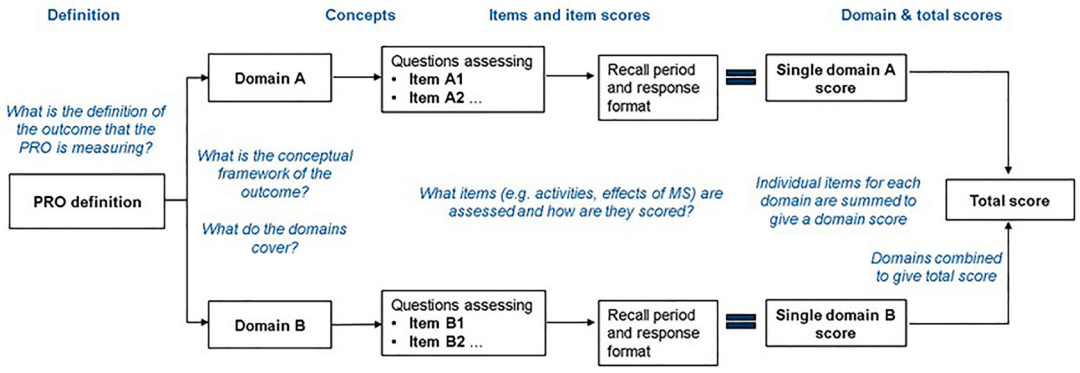

Figure 1. Flow-chart template for profiling the key elements of PRO instruments.* This standardized approach to profiling PROs allowed conclusions to be drawn about the extent to which scores generated by each instrument accurately reflect what, by definition, each PRO was designed to assess. *Note:* *Populated flow-charts for each of the six PROs are available in Supplementary Material
MS: multiple sclerosis; PRO: patient-reported outcome; VAS: visual analogue scale.

The aims of our research necessitated a mixed-methods approach. The quantitative investigation of the PROs allowed for a consistent assessment to be conducted across the PROs, while the qualitative work provided important insights from the patient perspective. Combining these methods enabled a more comprehensive examination of the subject than using either method in isolation.[19]

Results

Evaluation of PRO development

Table 2 shows data extracted for each PRO. **Supplementary Figure 1** shows each instrument's standardised flow chart. A detailed account of each instrument's development is available in the **supplementary material.** Below we provide a summary due to word count constraints.

Fatigue PROs

mFIS. The mFIS was developed in 1997[20] from the 40-item Fatigue Impact Scale (FIS).[11] The FIS developers aimed to develop a measure of PlwMS's perceptions of functional limitations attributable to fatigue. FIS items were selected from existing fatigue scales and $n = 30$ qualitative interviews with PlwMS. No explicit conceptual framework underpins the FIS. It was constructed to have 3 functioning subscales (physical $k = 10$ items; cognitive $k = 10$ items; psychosocial $k = 20$ items), reflecting the interview responses and dimensions from other health status and quality of life measures.[11] The 21-item mFIS was constructed by removing one item from the physical and 18 items from the psychosocial functioning dimensions. Neither the FIS nor mFIS developers define fatigue.

All FIS/mFIS items have five response categories from 0 (Never) to 4 (Almost always). The recall period is 1 month. mFIS/FIS items are summed to generate four scores: three subscale scores and a total score.

FSIQ-RMS. The FSIQ-RMS, developed in 2019,[10] aims to assess fatigue symptoms and their impacts on people with RMS. It has 20-items (7 symptoms, 13 impacts). The developers do not define explicitly what they mean by fatigue.

The FSIQ-RMS was developed in stages. A literature review led to preliminary conceptual frameworks for fatigue symptoms and impacts in RMS. Details are not given. These informed interview guides for $n =$ 17 concept elicitation interviews with PlwMS, resulting in the generation of 84 fatigue-related symptom and impact concepts. Concepts reported by ≥30% of interviewees were retained as a preliminary 30-item instrument, which was cognitively interviewed and completed by $n = 20$ PlwMS. Findings informed further item revisions, resulting in an instrument with 22 items (8 symptoms, 14 impacts).

Rasch analysis of the $n = 20$ completions of the preliminary 30-item scale led to one excluded item being re-introduced ($k = 23$; $k = 9$ symptoms; $k = 14$ impacts). This version was administered to PlwMS ($n = 164$) and controls ($n = 74$). Response data were analysed (floor/ceiling effects, item-item correlations, exploratory factor, Rasch analyses). The symptoms items were reduced from 9 to 7 due to redundancy.

Table 1. Characteristics of study participants.

Variable	*N* (%)
Sex	
Female	18 (72)
Male	7 (28)
Age	
21–30 years	6 (24)
31–40 years	9 (36)
41–50 years	5 (20)
51–60 years	3 (12)
61–70 years	2 (8)
Country	
France	1 (4)
Germany	3 (12)
Ireland	5 (20)
Italy	2 (8)
Luxembourg	1 (4)
Spain	1 (4)
UK	4 (16)
US	8 (32)
Ethnicity	
Asian	1 (4)
Black	2 (8)
Hispanic	1 (4)
Mixed	4 (16)
White	16 (64)
MS type	
RRMS	23 (92)
SPMS	2 (8)
Disease duration	
0–5 years	7 (28)
6–10 years	5 (20)
11 + years	13 (52)
Relationship status	
Divorced/separated	3 (12)
Married/co-habiting	15 (60)
Single	7 (28)

One impacts item was removed because of high ceiling effects. The exploratory factor analysis implied the remaining 13 impacts items exist in three 5-item subdomains (physical, cognitive and emotional, and coping).

FSIQ-RMS generates four scores: 1 symptom and 3 impacts subdomain scores.[10]

Life quality PROs

LMSQoL. The LMSQoL, developed in 2001,[14] is a patient-completed MS-specific quality of life measure. It has 8 items, each has four response categories (0 = Not at all to 3 = Most of the time).[21] The recall period is the past month. Item scores are summed to generate a total score. There was no *a priori* explicit definition of quality of life nor conceptual framework for measurement. It measures 'a variable related to well-being'.

Instrument development had multiple stages, including two focus groups with PlwMS ($n = 30$). The first identified 'the main areas of concern' and the second generated 25 potential items for instrument inclusion.[14] These were completed by $n = 24$ people. Analyses of response data (Cronbach's alpha and Rasch analyses) and consideration of item relevance led to 8 items being removed. The 17 remaining items had floor and ceiling effects, so 3 new items were identified and added to reflect the extremes of the QoL variable. The resultant 20-item scale was examined in two samples, one for test-retest reproducibility ($n = 27$) the other for construct validity ($n = 43$). Rasch analyses of the $n = 43$ completions identified 4 misfitting items, which were removed. The revised 16-item scale was administered to a stratified community sample. Rasch analysis of response data from $n = 180$ completions identified 8 misfitting items that were removed, leaving the final 8 item scale.

MSQoL-54. The MSQoL-54, developed in 1995, was created by adding 18 MS-specific items to the 36 generic items of RAND's 36-item health survey (SF-36).[22] The aim was to develop a measure of health-related quality of life (HRQoL) for MS combining the ability to compare across diseases and provide sensitive within-disease comparisons.

The SF-36 has 8 multi-item subscales (physical and social functioning, physical and emotional role limitations, general health perceptions, energy/vitality, emotional well-being, pain) and one single item assessing change in health over the last year. Of the 18 MS-specific items: 3 items are added to three existing SF-36 subscales (Pain, Energy, Social Functioning); 14 items are added as 4 new multi-item subscales (Health distress, cognitive and sexual functioning, QoL); one item assesses satisfaction with sex.[13]

The 18 MS items were generated by a literature review and input from specialist MS healthcare providers ($n = 3$), covering aspects understood to be particularly relevant to PlwMS. Recall periods and item response formats vary between subscales. No explicit conceptual framework or PlwMS involvement guided

Table 2. Data extraction summary by PRO instrument.

Instrument name	Focus of instrument	Target variable defined explicitly?[a]	Based on an a priori conceptual framework?	PlwMS involvement?	Domains and sub-domains assessed	Items measured[b]	Recall periods for items
LMSQoL[14,21c]	Measure QoL (well-being) specific to people with MS	No	No	Yes (Two focus groups to identify areas of concern and potential instrument items)	QoL (8 items)	*I have:* – Felt that my health has affected my relationships with my family – Felt lonely – Felt good about my appearance – Worried about my health – Worried about other people's attitudes about me – Felt tired – Had as much energy as usual – Felt happy about the future	**Past month** 4-point Likert scale, scored 0–3, ranging from *'Not at all'* to *'Most of the time'*, where a high score represents a worse QoL
MSQoL-54[13]	Measure HRQoL in people with MS	Yes: Health-related quality of life is described as a multidimensional construct that includes physical, mental and social health	No	No	Physical health (10 items)	*Does your health limit these activities; and by how much?* – Vigorous activities (such as running, lifting, sports)? – Moderate activities (such as moving a table, vacuuming)? – Lifting or carrying groceries? – Climbing one flight of stairs? – Climbing several flights of stairs? – Bending, kneeling or stooping? – Walking ≥1 mile? – Walking 1 block? – Walking several blocks? – Bathing and dressing yourself?	**In a typical day** 3-point scale, ranging from *'Yes, limited a lot'* to *'No, not limited at all'*
					Role limitations – physical problems (4 items)	*Has your health caused problems with work or activities:* – Limited the kind of work or activities? – Cut down time spent on work or other activities? – Accomplished less than you wanted? – Difficulty performing work or other activities?	**Past 4 weeks** Scored as '*Yes'* or '*No'*
					Role limitations – emotional problems (3 items)	*Has your health caused problems with work or activities:* – Cut down time spent on work or other activities? – Accomplished less than you wanted? – Didn't do work or other activities as carefully as usual?	**Past 4 weeks** Scored as '*Yes'* or '*No'*
					Pain (3 items)	*How much:* – Bodily pain have you been in? – Has pain interfered with your normal work? – Has pain interfered with your enjoyment of life?	**Past 4 weeks** Item 1: 6-point Likert scale, scored 1–6, ranging from *'None'* to *'Very severe'* Items 2–3: 5-point Likert scale, scored 1–5, ranging from *'Not at all'* to *'Extremely'*

(continued)

Table 2. Continued.

Instrument name	Focus of instrument	Target variable defined explicitly?[a]	Based on an a priori conceptual framework?	PlwMS involvement?	Domains and sub-domains assessed	Items measured[b]	Recall periods for items
					Emotional well-being (5 items)	*How much of the time have you:* – Been a very nervous person? – Felt so down in the dumps that nothing could cheer you up? – Felt calm and peaceful? – Felt downhearted and blue? – Been a happy person?	**Past 4 weeks** 6-point Likert scale, scored 1–6, ranging from *'All of the time'* to *'None of the time'*
					Energy/fatigue (5 items)	*How much of the time:* – Did you feel full of pep? – Did you have a lot of energy? – Did you feel worn out? – Did you feel tired? – Did you feel rested on waking in the morning?	**Past 4 weeks** 6-point Likert scale, scored 1–6, ranging from *'All of the time'* to *'None of the time'*
					Health perceptions (5 items)	– How is your health in general? *Rate how true or false the following statements are for you:* – I seem to get sick a little easier than other people – I am as healthy as anybody I know – I expect my health to get worse – My health is excellent	**In general** Item 1: 5-point Likert scale, scored 1–5, ranging from *'Excellent'* to *'Poor'* Items 2–5: 5-point Likert scale, scored 1–5, ranging from *'Definitely true'* to *'Definitely false'*
					Social function (3 items)	– To what extent has your physical health/emotional problems affected social activities? – How much time has health/ emotional problems affected social activities? – To what extent have problems with your bowel/bladder affected normal social activities?	**Past 4 weeks** 5-point Likert scale, scored 1–5, ranging from *'Not at all'* to *'Extremely'*
					Cognitive function (4 items)	*How much of the time:* – Have you had difficulty concentrating/thinking? – Did you have trouble keeping your attention on an activity for long? – Have you had trouble with your memory? – Have others noticed that you have trouble with memory/ concentration?	**Past 4 weeks** 6-point Likert scale, scored 1–6, ranging from *'All of the time'* to *'None of the time'*
					Health distress (4 items)	*How much of the time:* – Were you discouraged by your health problems? – Were you frustrated about your health? – Was your health a worry in your life? – Did you feel weighed down by your health problems?	**Past 4 weeks** 6-point Likert scale, scored 1–6, ranging from *'All of the time'* to *'None of the time'*
					Overall quality of life (2 items)	– Overall, how would you rate your own quality-of-life? – Which best describes how you feel about your life as a whole?	**In general** Item 1: VAS, ranging 0–10, whereby 10 is the *'Best possible quality of life'* and 0

(continued)

Table 2. Continued.

Instrument name	Focus of instrument	Target variable defined explicitly?[a]	Based on an a priori conceptual framework?	PlwMS involvement?	Domains and sub-domains assessed	Items measured[b]	Recall periods for items
							is the *'Worst possible quality of life'* Item 2: 7-point scale, scored 1–7, ranging from *'Terrible'* to *'Delighted'*
					Sexual function (4 items)	*How much of a problem was each of the following for you?* Men: – Lack of sexual interest? – Difficulty getting or keeping an erection? – Difficulty having orgasm? – Ability to satisfy sexual partner? Women: – Lack of sexual interest? – Inadequate lubrication? – Difficulty having orgasm? – Ability to satisfy sexual partner?	**Past 4 weeks** 4-point Likert scale, scored 1–4, ranging from *'Not a problem'* to *'Very much a problem'*
					Change in health (1 item)	– Compared to 1 year ago, how would you rate your health in general now?	**1 year** 5-point Likert scale, scored 1–5, ranging from *'Much better now than one year ago'* to *'Much worse now than one year ago'*
					Satisfaction with sexual function (1 item)	– Overall, how satisfied were you with your sexual function?	**Past 4 weeks** 5-point Likert scale, scored 1–5, ranging from *'Very satisfied'* to *'very dissatisfied'*
MSIS-29[16d]	Physical and psychological impact of MS	No	No	Yes (Semi-structured interviews to generate initial item pool)	Physical (20 items)	How much has MS limited your ability to: – Undertake physically demanding tasks? – Grip things tightly (e.g. turning on taps)? – Carry things?How much have you been bothered by: – Problems with your balance? – Difficulties moving about indoors? – Being clumsy? – Stiffness? – Heavy arms and/or legs? – Tremor of your arms or legs? – Spasms in your limbs? – Your body not doing what you want it to do? – Having to depend on others to do things for you? – Limitations in your social and leisure activities at home? – Being stuck at home more than you would like to be? – Difficulties using your hands in everyday tasks? – Having to cut down the amount of time you spent on work or other	**Past 2 weeks** 5-point Likert scale, scored 1–5, ranging from *'Not at all'* to *'Extremely'*

(continued)

Table 2. Continued.

Instrument name	Focus of instrument	Target variable defined explicitly?[a]	Based on an a priori conceptual framework?	PlwMS involvement?	Domains and sub-domains assessed	Items measured[b]	Recall periods for items
						– Problems using transport (e.g. car, bus, train, taxi, etc.)? – Taking longer to do things? – Difficulty doing things spontaneously (e.g. going out on the spur of the moment)? – Needing to go to the toilet urgently?	
					Psychological (9 items)	How much have you been bothered by: – Feeling unwell? – Problems sleeping? – Feeling mentally fatigued? – Worries relating to your MS? – Feeling anxious or tense? – Feeling irritable, impatient, or short tempered? – Problems concentrating? – Lack of confidence? – Feeling depressed?	**Past 2 weeks** 5-point Likert scale, scored 1–5, ranging from *'Not at all'* to *'Extremely'*
EQ-5D[23e]	Generic measure of health status	No	No	No	Mobility	See **Supplementary Figure 2** for item wording	**Measured 'Today'** 5-point Likert scale: choose the most appropriate description on the day
					Self-care	See **Supplementary Figure 2** for item wording	**Measured 'Today'** 5-point Likert scale: choose the most appropriate description on the day
					Usual activities	See **Supplementary Figure 2** for item wording	**Measured 'Today'** 5-point Likert scale: choose the most appropriate description on the day
					Pain/discomfort	See **Supplementary Figure 2** for item wording	**Measured 'Today'** 5-point Likert scale: choose the most appropriate description on the day
					Anxiety/depression	See **Supplementary Figure 2** for item wording	**Measured 'Today'** 5-point Likert scale: choose the most appropriate description on the day
					Health state	See **Supplementary Figure 2** for item wording	**Measured 'Today'** VAS, ranging 0–100, whereby 0 is the worst health you can imagine and 100 is the best health you can imagine
FSIQ-RMS[10f]	Measure fatigue symptoms and	No	Yes: fatigue related symptoms of	Yes (concept elicitation	Symptoms (7 items)	See **Supplementary Figure 3** for item wording	**Past 24 h** VAS, ranging 0–10,

(continued)

Table 2. Continued.

Instrument name	Focus of instrument	Target variable defined explicitly?[a]	Based on an a priori conceptual framework?	PlwMS involvement?	Domains and sub-domains assessed		Items measured[b]	Recall periods for items
	impacts in RMS		RMS and fatigue-related impacts of RMS based on literature search	interviews and cognitive interviews)				whereby 0 is *'Not at all'* and 10 is *'Extremely'*
					Impact (13 items)	Physical	See **Supplementary Figure 3** for item wording	**Past 7 days** 5-point Likert scale, scored 0–4, ranging from *'No difficulty/ Not difficult/Not at all/Never'* to 'Extreme difficulty/ Extremely difficult/ Extremely/Almost all of the time'
						Cognitive, emotional	See **Supplementary Figure 3** for item wording	**Past 7 days** 5-point Likert scale, scored 0–4, ranging from *'No difficulty/ Not difficult/Not at all/Never'* to 'Extreme difficulty/ Extremely difficult/ Extremely/Almost all of the time'
						Coping	See **Supplementary Figure 3** for item wording	**Past 7 days** 5-point Likert scale, scored 0–4, ranging from *'No difficulty/ Not difficult/Not at all/Never'* to 'Extreme difficulty/ Extremely difficult/ Extremely/Almost all of the time'
mFIS[20]	Measure PlwMS perceptions of the functional limitations that they attributed to their symptoms of fatigue	Yes: fatigue is described as a subjective lack of physical or mental energy that is perceived by the individual or caregiver to interfere with activities of daily living	No	Yes (original FIS developed using interviews with PlwMS)	Cognitive (10 items)		*Because of my fatigue,* – I have been less alert – I have difficulty paying attention for long periods of time – I have been unable to think clearly – I have been forgetful – I have difficulty making decisions – I have been less motivated to do anything that requires thinking – I have trouble finishing tasks that require thinking – I have difficulty organising thoughts – My thinking has slowed down – I have trouble concentrating	**Past 4 weeks** 5-point Likert scale, scored 0–4, ranging from *'Never'* to *'Almost always'*
					Physical (9 items)		*Because of my fatigue:* – I have been clumsy and uncoordinated – I have had to pace myself – I have been less motivated to do anything that requires physical effort – I have trouble maintaining activities for long periods of time – My muscles have felt weak	**Past 4 weeks** 5-point Likert scale, scored 0–4, ranging from *'Never'* to *'Almost always'*

(continued)

Table 2. Continued.

Instrument name	Focus of instrument	Target variable defined explicitly?[a]	Based on an a priori conceptual framework?	PlwMS involvement?	Domains and sub-domains assessed	Items measured[b]	Recall periods for items
						– I have been physically uncomfortable – I have been less able to complete tasks that require physical effort – I have limited my physical activities – I have needed to rest more often or for longer periods	
					Psychosocial (2 items)	*Because of my fatigue, I am:* – Less motivated to participate in social activities – Limited in my ability to do things away from home	**Past 4 weeks** 5-point Likert scale, scored 0–4, ranging from *'Never'* to *'Almost always'*

[a]In either the development publication or the instrument itself.
[b]Questions/items have been condensed for brevity; they are not intended to be comprehensive or verbatim.
[c]LMSQoL is a copyright of the University of Leeds, and is available from the University of Leeds fast-licence platform: https://licensing.leeds.ac.uk/product/lms-qol-leeds-multiple-sclerosis-quality-of-life-scale.
[d]MSIS-29 is a copyright of the University of Plymouth and is used under permission/licence.
[e]© EuroQol Research Foundation. EQ-5D™ is a trade mark of the EuroQol Research Foundation.
[f]FSIQ-RMS © 2017 Mapi Research Trust.
FSIQ-RMS: Fatigue Symptoms and Impacts Questionnaire – Relapsing Multiple Sclerosis; HRQoL: health-related QoL; LMSQoL: Leeds MS QoL instrument; mFIS: modified Fatigue Impact Scale; MS: multiple sclerosis; MSIS-29: 29-item MS Impact Scale; PlwMS: people living with MS; PRO: patient-reported outcome; QoL: quality of life; RMS: relapsing multiple sclerosis.

MSQoL-54 development. Measurement performance testing was conducted in 179 PlwMS.[13]

EQ-5D. The EQ-5D, first published in 1990,[23] was developed as a standardised non-disease-specific instrument for describing and valuing HRQoL. There are three-level and five level versions (EQ-5D-3L; EQ-5D-5L), differing only in the number of response categories of the five items (dimensions). Both versions have two parts. First, 5 items grading severity of problems with mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Second, a visual analogue scale (VAS) rating health state from 0 = worst to 100 = best imaginable. In the EQ-5D-3L, each item has 3 severity levels (no problems; some/moderate; extreme problems/unable to do); the 5-level version has 5 severity (none; slight, moderate, severe, unable/extreme).

The developers do not give definitions of health status or quality of life, and there is no explicit underpinning conceptual framework. Whilst the items were selected from a 'detailed examination of the descriptive content of existing health status measures',[23] and there was 'agreement among scientists and clinicians', specific detail is not documented. Details of patient involvement have not been published.

EQ-5Ds generate three scores: an unweighted health status profile across the five dimensions (e.g. 1 2 1 2 3), a weighted sum score (index) derived from the unweighted profile, and a VAS health state score.[15,24,25] The recall period is 'today'.

MS impact

MSIS-29. The MSIS-29 was developed in 1999 to measure the physical (20-items) and psychological (9-items) impact of MS. Two scores are generated, one for each domain. All items in MSIS-29v2 have 4 response categories.[26]

There was no definition of MS impact reported. No *a priori* conceptual framework underpinned MSIS-29 development. Development involved input from multiple disciplines, expert opinions, literature review, and PlwMS. A pool of $k = 141$ items was generated from $n = 30$ semi-structured interviews with PlwMS. These were administered to a random sample of $n = 1530$ PlwMS. The $k = 129$ non-walking-related items were analysed using traditional psychometric methods and then reduced using statistical criteria to form the two subscales.[27] The item content of the scales drove the name of their measurement variables

and these were refined accordingly. MSIS-29's measurement performance was determined from an independent sample of $n = 1250$ PlwMS, test-retest reproducibility in a subgroup of these and responsiveness in an independent sample of $n = 55$. The 12 walking-related items formed the MSWS-12.[28]

Interview insights on PROs from PlwMS ($n = 22$)

Fatigue PROs

mFIS. PlwMS reported that the mFIS provided an accurate description of fatigue, the questions related to cognition and fatigue were relevant, and that the scale was clear. PlwMS commented that the mFIS does not assess the impact of fatigue on emotions or other aspects of everyday life. The scoring was described as confusing as questions are both positively and negatively phrased. Furthermore, PlwMS suggested a 4-week recall period is too short to measure the impact of fatigue on everyday life (Table 3). Respondents did not suggest alternative timeframes.

FSIQ-RMS. PlwMS felt the FSIQ-RMS covered a wide range of domains. The assessment of symptoms and functional impacts was welcomed. Digital administration was deemed convenient. Perceived weaknesses included the limited timeframe (24 h for symptoms, 7 days for impacts), the questionnaire was considered too long (20 items), and psychosocial aspects were not covered comprehensively (Table 3).

Life quality PROs

LMSQoL. PlwMS said the LMSQoL contained thought-provoking questions, reflected the connection between mental health and MS, and was a good instrument to track changes in symptoms. However, interviewees said the LMSQoL did not adequately reflect the relationship between physical and emotional symptoms or the relationship between fatigue, and cognitive and sexual functioning (Table 3).

MSQoL-54. PlwMS said the MSQoL-54 questions provide a holistic view of their MS experience, covered most of the emotional aspects of MS, provided a variety of response formats, and considered the fluctuating nature of their symptoms. However, PlwMS felt the weaknesses were the PRO's length (54 items), focus on disability rather than ability, and limited range of pain questions. They recommended the wording of sex-related questions could be more considered (less direct) (Table 3).

EQ-5D. Interviewees appreciated the EQ-5D was short, simple, quick to complete, and covered relevant topics pertaining to general health. However, they reported that the questions were not detailed enough, overly simple or too generic (while recognising that this is a general non-MS specific measure). They found the five-digit summary score hard to relate to **(**Table 3**)**.

MS impact PRO

MSIS-29. PlwMS said the MSIS-29 included relatable question wording, covered a diverse range of topics, and explored both physical and psychological impacts of MS. However, interviewees highlighted the unequal focus on the two domains, with fewer items dedicated to psychological aspects. Respondents also reported an insufficient focus on pain.

Table 4 summarises key interview insights. Various themes emerged. Patients reported that there is no one-size fits all PRO, and that it would be helpful for instruments to be tailored to specific relevant characteristics like disease type/stage or cultural background. It was also recognised that different people prefer different modes of instrument completion (e.g. paper and pencil or digital) and that these preferences should be accounted for during administration. With regards to fatigue, respondents stated that instruments often do not adequately capture its impact, especially the fluctuating nature of this symptom. The interviewees also expressed that it is important for them to fully understand the purpose of the PRO instrument and how it will support the delivery of optimal care. Other key themes to emerge included the need for questions to be simple, carefully worded, and relevant, for response scales to be meaningfully related to the symptom in question, and that recall periods account for potential memory impairment.

Survey insights on PROs from PlwMS

Figure 2 shows the results of the post-interview interviewee survey on topics related to the six PROs. For the fatigue-specific PROs, participants generally agreed that the two PROs effectively assessed the impact of MS on their fatigue, covered aspects relevant to PlwMS, weren't significantly time-demanding, and would be manageable to complete. In relation to the other PROs, a number of respondents said the LMSQoL ($n = 7$, 39%) and EQ-5D ($n = 8$, 44%) could not effectively assess the impact of MS on their life quality, or that they covered aspects relevant to them as a PlwMS. Most respondents said all PROs

Table 3. PlwMS feedback of PROs.

PlwMS feedback on mFIS		
Strengths	Weaknesses	Suggested improvements
– Good psychosocial assessment – Scale is clear and relevant – Accurate description on the scale of fatigue – Cognition questions are relevant – Fatigue questions are relevant	– Only measures over a 4-week recall period – Lacks recognition of an emotional impact – Lacks recognition of impact on everyday life – Scoring can be confusing (depending on the way the question is either positively or negatively phrased, the scoring is inversed)	– Inclusion of more psychosocial questions – Rewording of questions to lay language – Simplify scoring
PlwMS feedback on FSIQ-RMS		
Strengths	**Weaknesses**	**Suggested improvements**
– Broad range of questions covering subjects relevant to PlwMS – Focuses on practical situations – Measures coping with MS symptoms – Includes cognitive, physical and psychosocial elements – The instrument is simple whilst reaching a good level of detail – Easy digital access – Explores the impact of each symptom presented	– Only covers a recall period of 24 h and impact for 7 days – Length of the PRO may be burdensome – Psychosocial questions are not comprehensive enough	– Increase recall period
PlwMS feedback on MSQoL-54		
Strengths	**Weaknesses**	**Suggested improvements**
– Questions provide a holistic view of the PlwMS's experience of MS – Questions address most of the emotional aspects – The wide spectrum of symptoms demonstrates an understanding of the PlwMS' reality – Answers are not restricted to a set scale – The instrument considers fluctuations in MS symptoms	– The scale scores are not well described and have gaps (particularly for recall time of symptoms) – Focuses too much on what PlwMS cannot do rather on what they can do – Lack of exploration around pain – Wording of questions is hard to relate to – Length of the PRO may be burdensome – Addressing matters of sexual function needs less direct/considered wording	– Update the language to a more modern and relatable style – Questions to be phrased more positively – Update the questions to reflect more recent science and how patients live with MS in today's world
PlwMS feedback on LMSQoL		
Strengths	**Weaknesses**	**Suggested improvements**
– Good choice of questions – Contains detailed questions	– The relationship between the	– Remove the question relating to

(continued)

Table 3. Continued.

PlwMS feedback on mFIS		
Strengths	Weaknesses	Suggested improvements
that can be informative and thought provoking for PlwMS – Makes the connection between mental health issues and MS – Good instrument to track changes in MS symptoms	physical and emotional symptoms of MS is not addressed – The relationship between fatigue and cognitive or sexual function is not addressed	appearance ("I have felt good about my appearance") – Use a different scoring scale – Many questions in this PRO would benefit from a follow-up discussion with a health care professional
PlwMS feedback of EQ-5D		
Strengths	**Weaknesses**	**Suggested improvements**
– Covers relevant topics about general health (covers the basics) – The tool is quick, short and simple	– Instrument is not MS specific – Not very detailed and overly simplified – 5-digit number system hard to relate to – Items are sometimes perceived as too generic – Does not address cognitive function	– The mobility questions do not reflect the realities of PlwMS – Add an introduction relating to the purpose/aims of the instrument
PlwMS feedback of MSIS-29		
Strengths	**Weaknesses**	**Suggested improvements**
– Questions worded in a relatable style – Covers a diverse range of relevant topics – Explores not just the physical but also the psychological impact – Good level of detail	– Not enough focus on psychological impacts compared with physical impacts – The items relating to physically demanding tasks are described too vaguely – Does not sufficiently address pain – Does not measure impact of MS on daily life	– Clearly describe the impact of MS on the items being measured

FSIQ-RMS: Fatigue Symptoms and Impacts Questionnaire – Relapsing Multiple Sclerosis; HRQoL: health-related QoL; LMSQoL: Leeds MS QoL instrument; mFIS: modified Fatigue Impact Scale; MS: multiple sclerosis; MSIS-29: 29-item MS Impact Scale; PlwMS: people living with MS; PRO: patient reported outcome; QoL: quality of life.

would be manageable to complete. Five respondents (28%) said the MSQoL-54 would create a significant time burden to complete.

Discussion

When PROs are used in MS studies we assume the measured effect adequately approximates the actual effect. More specifically, we assume the PROs provide accurate and precise measurements of the clinical variables they purport to measure, and changes in PRO scores are valid indications of what happens in practice. For example, if fatigue is measured with a PRO in a treatment trial, and data show fatigue is not improved, we assume this reflects what happens in practice. These assumptions are requirements for high stakes MS studies (i.e. pivotal trials of MS disease modifying therapies that ultimately have implications for the care of individuals and the expenditure of public funds). It is noteworthy that suboptimal measurement generates type 2 errors.[7]

This study concerns one of these requirements: how we can begin to tell if PROs generate valid measurements. We examined the development of selected PROs rather than their psychometric (statistical) performance, because PRO validity is primarily determined by its structure and item content more than its psychometric performance. The limitations of

Table 4. Summary of key insights from PlwMS on PROs.

Theme	Key insights
Individuality	– There is no one-size fits all PRO. Individuality is multi-stranded; the personality and background of the PlwMS play an important role in coping with MS and the resulting perceptions of how the disease changes their life and physiology.
Personalisation	– PROs should be tailored to the stage/type of MS. – The geographical and cultural background of PlwMS should be taken into consideration.
Choice	– PlwMS can be empowered to participate in PROs by offering a choice of administration style (e.g. audio recording, digital, paper-based, face to face interview style) and in turn, this may lead to greater levels of insight. – Different PlwMS like different ways of answering questions, with answers ranging from a preference for scaling to a preference for interview-style reporting of symptoms. – PlwMS would like the choice of using PROs to measure changes over time in conjunction with routine clinical practice, as well as in clinical trials. – The ability to choose when to complete a PRO (e.g. before coming into the clinical setting) could avoid stress and improve the quality of answers.
Communication	– Relatability is key: patients stated that the style of questions are not formulated with enough specificity. – PlwMS can feel misunderstood, especially when explaining the impact of living with fatigue; often not adequately captured by PROs, nor do PROs take into account the short- and long-term fluctuations of fatigue. – Greater psychoeducational support is required to help patients learn how to communicate their fatigue, and campaigns are needed to develop a greater awareness of cognitive impairments triggered either by MS or co-existing fatigue or depression.
Clarity	– PlwMS need to understand the purpose and importance of PROs and how they support the delivery of optimal care.
Language and terminology	– Careful wording of the questions is essential to generate valid and meaningful responses. – PlwMS appreciate simplicity in communication but the wording needs to find the right balance between an overcomplicating and patronising tone.
Scaling	– PlwMS require symptom scales that reflect the experience of the symptom in a way that is meaningful to them.
Recall period	– There are mixed views on the right length of *recall* (from '24 h ago', 'a week ago', or 'a month to a year ago'). Factors such as fatigue, cognition and mood at the time of recall may play a role. Additionally, MS symptoms fluctuate and the phrasing of the recall-based questions should reflect this.
Autonomous tracking	– PlwMS feel empowered by being able to record changes in their illness and use different methods of logging their symptoms (e.g. keeping a diary, making lists, using digital application).
Emotional impact	– The emotional impact of MS intrinsically runs throughout all other feedback and highlights how aspects such as anxiety, depression, pain and cognitive impairment are intricately linked.

MS: multiple sclerosis; PlwMS: people living with MS; PRO: patient-reported outcome.

statistical validity tests were highlighted decades ago,[29] and their potential to mislead demonstrated empirically more recently.[8]

We examined selected PROs for definitions of the variables they measured and their conceptual underpinnings. No development paper of any selected

PROs reviewed provided a thorough definition of the variables they sought to measure. Most gave no definition. Some determined the variable measured *post hoc* from the items left by statistically-driven reductions of larger item sets, or from correlations with other PROs. We found similar omissions of conceptual frameworks. Whilst all development papers provided, in differing forms, information on the structure of the final instrument *post hoc*, none provided conceptual frameworks *a priori* for the variable for measurement that enabled us to determine the validity of the final instrument. Most notable was the paucity of documented information about these essential aspects of scale development; without definitions and conceptualisation the extent to which scores reflect concepts is unknown.

Approaches to selected PRO development varied, yet all provided scores purporting to be valid measurements of clinical variables. Different criteria were applied. Many appeared arbitrary. Interestingly, a number of the PROs examined generated large amounts of information (concepts and items) from qualitative research, the majority of which was ultimately discarded. On no occasion was this rich information structured into an explicit framework to aid understanding of the variable of interest

We recognise our selection of PROs was limited and open to criticism. However, the issues we highlight are widespread in health measurement. For example, the widely used Fatigue Severity Scale (FSS) suffers from exactly the same weaknesses as the other fatigue PROs we have examined. The original development paper of 1989 does not include any definition of fatigue.[30] There is no conceptual underpinning. The 9 items were selected from a pool of 28 items (that was neither provided, described, nor referenced) based on 'a factor analysis, item analysis, and theoretic considerations'. As such, based on the original development paper for this instrument, we are left uncertain as to the extent to which FSS scores accurately reflect fatigue in MS.

It is curious that PROs have been developed to measure complex clinic variables, like fatigue and quality of life, without more attention to variable definitions and conceptualisations. There are a number of possible explanations. These include the absence of regulatory requirement until recently, limited guidance on defining and conceptualising variables, and scale development strategies dominated by quantitative methods with scant attention to qualitative techniques. Certainly, the scale development zeitgeist previously was a triad of generation of an item pool, scale formation by statistically driven item reduction, and statistical evaluation of the scales produced. It is hardly surprising that a set of items selected because they are statistically cohesive, are statistically cohesive when examined subsequently.

We also examined the roles of PlwMS in selected PRO development and gained PlwMS's feedback on the PROs. PlwMS's involvement in PRO development was varied from none to quite extensive. For some PROs it was unclear. Their feedback also varied. There were positive and negative comments for each PRO. Most PROs were considered to lack relevant components. We think the absence of variable definitions and conceptual frameworks makes it very difficult for PlwMS to critique PROs meaning. They need this information to set a frame of reference for their input. In addition, there is little guidance as to exactly how PlwMS, or other conditions, should best be involved in PRO development and evaluation, and how to maximise quality control of this process. We identified several key themes from the interviews that may serve to optimise the development of future PRO instruments. Some of these insights may be more challenging than others to incorporate, such as tailoring instruments to specific patient characteristics or using a recall period that is acceptable to everyone. However, many of these important interview insights are relatively straightforward and can be easily accommodated, such as providing a clear explanation of the PROs purpose, offering a choice of administration formats, using clear and simple question wording, and making sure that item wording reflects constructs that are meaningful and relevant to those who are responding. Implementing these insights to inform instrument development will ensure that future PROs are fit for purpose and acceptable to patients.

Findings from our interviews identified geographical and cultural background as important aspects to consider when developing PRO instruments. It is therefore critical that during the development of PRO instruments, PlwMS are not only heavily involved in the process, but that those involved represent a diverse range of characteristics, ensuring that the resultant instruments are fit for purpose across the widely diverse population of PlwMS. If not, then there is a risk that these tools may have limited or imperfect generalizability to under-represented minority groups. Interestingly, only two of the development papers for the six PRO instruments assessed in this study provided details on the ethnicity of the development sample. For the FSIQ-RMS the percentage of non-Caucasian/White participants ranged from 15% to 60% across the three

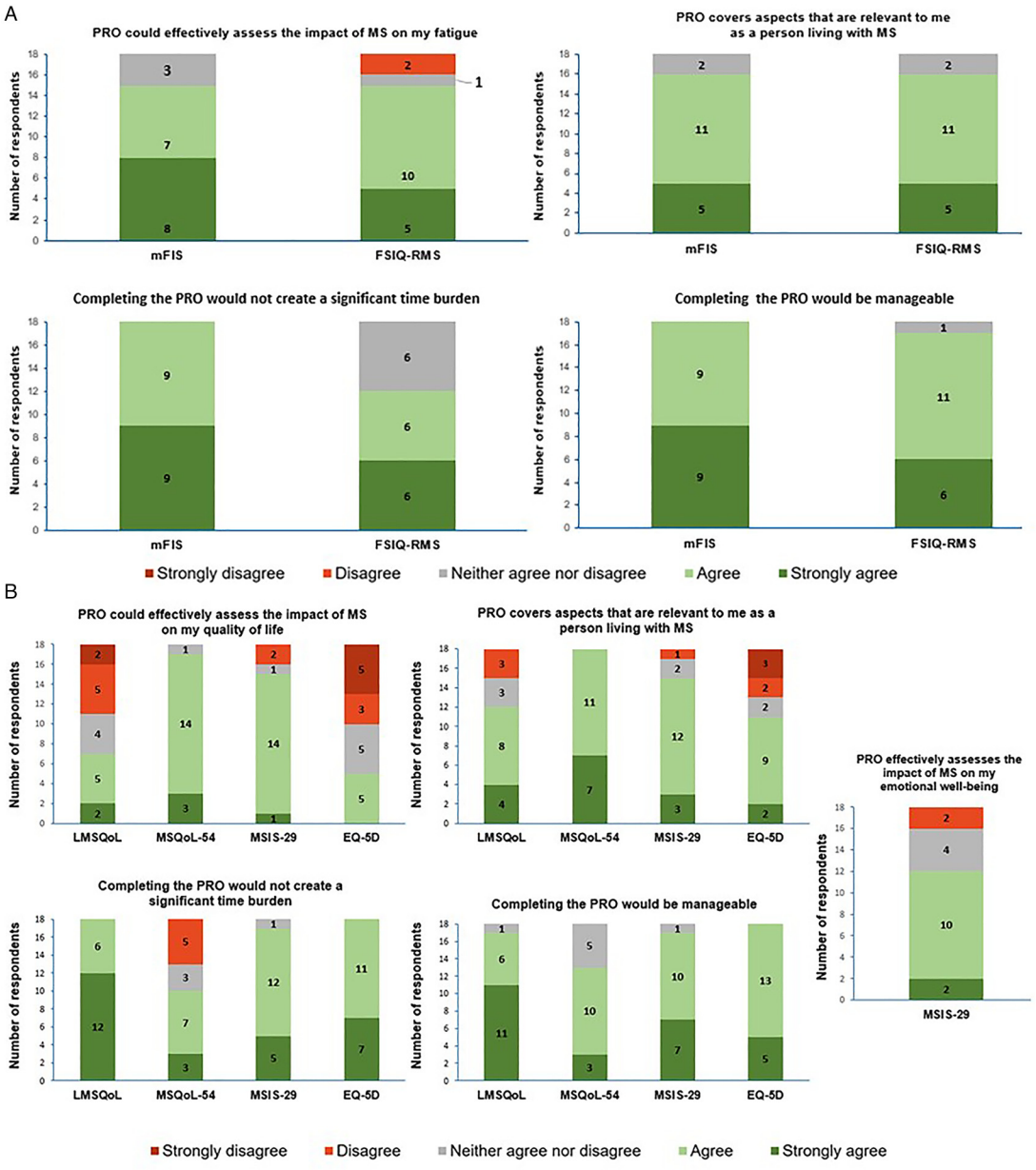

Figure 2. PlwMS post-interview survey on fatigue (A) and QoL or physical/psychological (B) PROs using a 5-point Likert scale. *Note*: Data represents responses from 18/22 interviewees. FSIQ-RMS: Fatigue Symptoms and Impacts Questionnaire – Relapsing Multiple Sclerosis; LMSQoL: Leeds MS QoL instrument; mFIS: modified Fatigue Impact Scale; MS: multiple sclerosis; MSQoL-54: 54-item MS QoL; MSIS-29: 29-item MS Impact Scale; PlwMS: people living with MS; PRO: patient-reported outcome; QoL: quality of life.

content development stages.[10] For the MSIS-29, the development sample was entirely white.[16]

Our findings reiterate the requirement for PRO developers to provide explicit definitions and detailed conceptualisation of the variables they seek to measure, as well as the importance of patient involvement. Whilst these are recognised requirements,[4,31,32] it appears that such guidance has generally not been followed to date. Future instrument development aligned to best practiced principles will result in fit-for-purpose PROs, enabling strategic PRO selection to underpin clinical-decisions in the care of PlwMS. Until then the validity of PRO data in MS remains questionable.

One reviewer of this study raised three very relevant and important questions. How should we interpret

studies using these instruments? What guidance is there for instrument selection? How do we optimize existing PROs while waiting for better instruments to be developed? Each question warrants a detailed answer beyond the scope of this manuscript and are being addressed actively by us in other studies. In short, it is difficult to quantify accurately the impact of poor measurement especially in the absence of clear definitions and conceptualisations of the clinical variables they seek to measure. Without that information the extent to which an instrument's score reflects the construct of interest, its validity, is unclear. Detailed head-to-head comparisons of PROs, using qualitative and sophisticated quantitative methods are required, in specific contexts of use, to enable clear understandings of the trade-offs associated with competing PROs. Such detailed evaluations enable PRO strengths and limitations to be identified. Importantly, they also act as a platform for PRO modification to maximise their current performance as measures. PRO measurement strategies must be well thought-out, critically appraised, and the associated science conducted robustly.

Acknowledgements

The authors would like to thank the people living with MS, who consented to be interviewed and completed our survey, for contributing their views and experiences to the PROMPT-MS initiative. We would also like to thank Nanette Rombach-Mullan and David McMinn, PhD (both of Novartis CONEXTS, Ireland) for providing medical writing support, which was funded by Novartis Pharma AG.

Declaration of conflicting interests

The authors declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article. Birgit Bauer has received compensation for consulting from Novartis, Roche, Merck, Teva and Sanofi. Disclosures do not show a conflict with the work being presented. Trishna Bharadia in the last 3 years has received compensation for serving as a consultant, writer and/or speaker for or has received honoraria from: 67Health, Abbvie, Actelion (Janssen), Admedicum, Blue Latitude Health (Fishawack), Corex Logistics, Curatio, DHL Life Sciences, Envision Pharma, Faculty of Pharmaceutical Medicine, Future Medicine, Gilead Sciences, Greenphire, ISMPP, Kayentis, Medipace, Merck KgA, NIHR, Norgine, Novartis, NovoNordisk, Parexel, Pfizer, Prime Global Roche, Savvy Cooperative, Synchrogenix (Certara), talkHealth, Teva, University College London, University of Central Lancashire, University of Surrey, WEGO Health, Wellcome Trust, and Vitaccess. Disclosures do not show a conflict with the work being presented. Tanuja Chitnis has received compensation for consulting from Biogen, Novartis Pharmaceuticals, Roche Genentech, and Sanofi Genzyme. She has received research support from the National Institutes of Health, National MS Society, US Department of Defense, Sumaira Foundation, Brainstorm Cell Therapeutics, EMD Serono, I-Mab Biopharma, Mallinckrodt ARD, Novartis Pharmaceuticals, Octave Bioscience, Roche Genentech, and Tiziana Life Sciences. Disclosures do not conflict with the work being presented. Piet Eelen has received compensation for consulting, advising and presenting from Merck, Convatec, Novartis and Biogen. Disclosures do not conflict with the actual work being presented. Giampaolo Brichetto has been member of the advisory board of Novartis and Roche. Disclosures do not conflict with the work being presented. Andrew Lloyd works for and holds stock in Acaster Lloyd Consulting Ltd which has received fees from Novartis. Disclosures do not show a conflict with the work being presented. Hollie Schmidt has received compensation for consulting from Celgene, and Accelerated Cure Project has received grants, collaboration funding and consulting payments from Biogen, Bristol Myers Squibb, Celgene, EMD Serono, Genentech, MedDay, Novartis and Sanofi Genzyme. Disclosures do not show a conflict with the work being presented. Jeremy Hobart has received consulting fees, honoraria, support to attend meetings or research support from Acorda, Asubio, Bayer Schering, Biogen Idec, F. Hoffmann-La Roche, Genzyme, Merck Serono, Novartis, Oxford PharmaGenesis and Teva. Disclosures do not show a conflict with the work being presented. Miriam King, Jennifer Fitzgerald, Thomas Hach, and Jo Vandercappellen, are employees of Novartis Pharma AG.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article. This work was supported by Novartis Pharma AG.

ORCID iDs

Trishna Bharadia https://orcid.org/0000-0003-3633-729X
Tanuja Chitnis https://orcid.org/0000-0002-9897-4422
Giampaolo Brichetto https://orcid.org/0000-0003-2026-3572

Supplemental material

Supplemental material for this article is available online.

References

1. Newland PK, Naismith RT and Ullione M. The impact of pain and other symptoms on quality of life in women with relapsing-remitting multiple sclerosis. *J Neurosci Nurs* 2009; 41: 322–328.
2. Janardhan V and Bakshi R. Quality of life in patients with multiple sclerosis: the impact of fatigue and depression. *J Neurol Sci* 2002; 205: 51–58.
3. Manjaly ZM, Harrison NA, Critchley HD, et al. Pathophysiological and cognitive mechanisms of fatigue in multiple sclerosis. *J Neurol Neurosurg Psychiatry* 2019; 90: 642–651.
4. FDA. FDA Guidance for Industry, http://www.fda.gov/downloads/Drugs/Guidances/UCM193282.pdf (2009).

5. Nowinski CJ, Miller DM and Cella D. Evolution of patient-reported outcomes and their role in multiple sclerosis clinical trials. *Neurotherapeutics* 2017; 14: 934–944.
6. Walton MK, Powers JH3rd, 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.
7. Hobart JC, Cano SJ, Zajicek JP, et al. Rating scales as outcome measures for clinical trials in neurology: problems, solutions, and recommendations. *Lancet Neurol* 2007; 6: 1094–1105.
8. Hobart J, Cano S, Baron R, et al. Achieving valid patient-reported outcomes measurement: a lesson from fatigue in multiple sclerosis. *Mult Scler* 2013; 19: 1773–1783.
9. Khurana V, Sharma H, Afroz N, et al. Patient-reported outcomes in multiple sclerosis: a systematic comparison of available measures. *Eur J Neurol* 2017; 24: 1099–1107.
10. Hudgens S, Schuler R, Stokes J, et al. Development and validation of the FSIQ-RMS: a new patient-reported questionnaire to assess symptoms and impacts of fatigue in relapsing multiple sclerosis. *Value Health* 2019; 22: 453–466.
11. 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: S79–S83.
12. Larson RD. Psychometric properties of the modified fatigue impact scale. *Int J MS Care* 2013; 15: 15–20.
13. Vickrey BG, Hays RD, Harooni R, et al. A health-related quality of life measure for multiple sclerosis. *Qual Life Res* 1995; 4: 187–206.
14. Ford HL, Gerry E, Tennant A, et al. Developing a disease-specific quality of life measure for people with multiple sclerosis. *Clin Rehabil* 2001; 15: 247–258.
15. Rabin R and de Charro F. EQ-5D: a measure of health status from the EuroQol group. *Ann Med* 2001; 33: 337–343.
16. Hobart J, Lamping D, Fitzpatrick R, et al. The multiple sclerosis impact scale (MSIS-29): a new patient-based outcome measure. *Brain* 2001; 124: 962–973.
17. EUPATI. European Patients' Academy on Therapeutic Innovation: Patient Engagement Through Education., https://eupati.eu/ (01 March 2021).
18. EphMRA guidelines https://www.ephmra.org/standards/code-of-conduct-aer/ (2020).
19. Halcomb E and Hickman L. Mixed methods research. *Nurs Stand* 2015; 29: 41–47. DOI: 10.7748/ns.29.32.41.e8858.
20. The Consortium of Multiple Sclerosis Centers Health Services Research Subcommittee. Multiple Sclerosis Quality of Life Inventory: A User's Manual. © National Multiple Sclerosis Society, 1997. https://www.nationalmssociety.org/NationalMSSociety/media/MSNationalFiles/Brochures/MSQLI_-A-User-s-Manual.pdf.
21. Ensari I, Motl RW and McAuley E. Structural and construct validity of the leeds multiple sclerosis quality of life scale. *Qual Life Res* 2016; 25: 1605–1611.
22. Hays RD, Sherbourne CD and Mazel RM. The RAND 36-item health survey 1.0. *Health Econ* 1993; 2: 217–227.
23. EuroQol G. EuroQol--a new facility for the measurement of health-related quality of life. *Health Policy* 1990; 16: 199–208.
24. Rabin R, Gudex C, Selai C, et al. From translation to version management: a history and review of methods for the cultural adaptation of the EuroQol five-dimensional questionnaire. *Value Health* 2014; 17: 70–76.
25. Lloyd A and Pickard AS. The EQ-5D and the EuroQol group. *Value Health* 2019; 22: 21–22.
26. Hobart J and Cano S. Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods. *Health Technol Assess* 2009; 13, iii, ix-x, 1–177.
27. Hobart JC, Riazi A, Lamping DL, et al. Improving the evaluation of therapeutic interventions in multiple sclerosis: development of a patient-based measure of outcome. *Health Technol Assess* 2004; 8: 1–48.
28. Hobart JC, Riazi A, Lamping DL, et al. Measuring the impact of MS on walking ability: the 12-item MS walking scale (MSWS-12). *Neurology* 2003; 60: 31–36.
29. Stenner AJ, Smith M and Burdick D. Towards a theory of construct definition. *J Educ Meas* 1983; 20: 305–316.
30. Krupp LB, LaRocca NG, Muir-Nash J, et al. The fatigue severity scale. Application to patients with multiple sclerosis and systemic lupus erythematosus. *Arch Neurol* 1989; 46: 1121–1123.
31. Terwee CB, Prinsen CAC, Chiarotto A, et al. COSMIN Methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Qual Life Res* 2018; 27: 1159–1170.
32. Carlton J, Peasgood T, Khan S, et al. An emerging framework for fully incorporating public involvement (PI) into patient-reported outcome measures (PROMs). *J Patient Rep Outcomes* 2020; 4. DOI: 10.1186/s41687-019-0172-8.