A PROMIS fatigue short form for use by individuals who have multiple sclerosis

Karon F. Cook,
Northwestern University, Chicago, IL, USA

Alyssa M. Bamer,
University of Washington, Seattle, WA, USA

Toni S. Roddey,
Texas Woman's University, Houston, TX, USA

George H. Kraft,
University of Washington, Seattle, WA, USA

Jiseon Kim, and
University of Washington, Seattle, WA, USA

Dagmar Amtmann
University of Washington, Seattle, WA, USA

Abstract

Purpose—To derive from the Patient Reported Outcomes Measurement Information System (PROMIS) fatigue item bank, a short form for individuals with multiple sclerosis (MS), the PROMIS-Fatigue$_{MS}$.

Methods—A panel of 37 clinicians and 46 individuals with MS ranked the relevance of PROMIS fatigue items to persons with MS. Eight items were selected for the PROMIS-Fatigue$_{MS}$ that maximized relevance rankings, content coverage, and item discrimination. The PROMIS-Fatigue$_{MS}$ and an existing, 7-item PROMIS fatigue short form (PROMIS-Fatigue$_{SFv1.0}$) were administered to a new sample of 231 individuals with MS. Known groups and content validity were assessed.

Results—Scores from the short forms were highly correlated ($r = 0.92$). Discriminant validity of the PROMIS-Fatigue$_{MS}$ scores was supported in known groups comparisons. Scores of neither short form exhibited an advantage in quantitative analyses. The PROMIS-Fatigue$_{MS}$ targeted more of the content included in participants' responses to open-ended questions than did the PROMIS-Fatigue$_{SFv1.0}$.

Conclusions—The PROMIS-Fatigue$_{MS}$ was derived to have content validity in MS samples. The validity of the measure was further supported by the ability of PROMIS-Fatigue$_{MS}$ items to discriminate among groups expected to differ in levels of fatigue. We recommend its use in measuring the fatigue of individuals with MS.

Correspondence to: Karon F. Cook.

Keywords

Introduction

Multiple sclerosis (MS) is characterized by numerous symptoms including physical, cognitive, and emotional problems, numbness, gait problems, bowel and bladder dysfunction, vision problems, dizziness and vertigo, sexual dysfunction, pain, depression, spasticity, and fatigue [1]. Fatigue is among the most common and debilitating of these symptoms, affecting approximately 80% of persons who have the disease [2–5]. In a community-based survey, more than 83% of people with MS reported experiencing fatigue [2], and 69% considered fatigue their worst symptom [6]. Fatigue in MS may directly impact participation in important roles such as employment [7, 8] and can profoundly magnify other MS symptoms [9]. Currently, there is no laboratory test to measure fatigue; therefore, its assessment is typically accomplished through self-report.

In recent years, there has been a shift to the use of modern psychometric approaches such as item response theory (IRT) and item banking in developing self-report health outcome measures. Once an item bank is developed, items can be administered adaptively (computer adaptive testing), and short forms can be derived from the bank to target particular populations or assessment contexts [10]. An example of a recent item banking effort is the Patient Reported Outcome Measurement Information System (PROMIS) [11]. The PROMIS initiative was funded by the National Institutes of Health to develop item banks to measure symptoms and quality-of-life indicators applicable to a range of chronic conditions. A unique characteristic of the scores of PROMIS measures is that scores are reported on a T-score metric that is anchored to mean score levels in a healthy U.S. general population [12]. The T-score metric has a mean of 50 and a standard deviation of 10. The use of this metric improves the interpretability of scores. For example, a score of 60 on a PROMIS measure is one standard deviation above the mean of a healthy normative U.S. sample.

Among the PROMIS, item banks is one developed to measure fatigue, defined as "an overwhelming, debilitating, and sustained sense of exhaustion that decreases one's ability to carry out daily activities, including the ability to work effectively and to function at one's usual level in family or social roles" (p. 1318) [13]. Three published short forms have been derived from the PROMIS Fatigue bank [14]. The 7 items of Version 1 (PROMIS-Fatigue$_{SFv1.0}$) were chosen to represent the range of content in the PROMIS Fatigue item bank. Thus, this short form is an "all-purpose" measure designed for use across chronic diseases. However, some have argued that the content validity of such measures should be evaluated in each clinical population in which the measure may be used [15].

This study had two purposes. The first was to derive from the PROMIS fatigue item bank a new short form, the PROMIS-Fatigue$_{MS}$, comprised of items relevant to individuals with MS. The second objective was to administer the PROMIS-Fatigue$_{MS}$ and the PROMIS-Fatigue$_{SFv1.0}$ in a new sample of individual with MS and compare the known groups and content validity of scores on the two short forms.

Methods

Institution Review Board approval was obtained for each part of the study involving individuals with MS.

Objective 1: identify items for the PROMIS-Fatigue$_{MS}$ relevant to the experience of fatigue in MS

Expert panel and participants with MS—To help identify a subset of items relevant to the experience of fatigue in MS, we recruited an expert panel of clinicians with experience treating MS and a sample of individuals living with MS. The clinician expert panel was recruited from among: (1) MS physicians from the Departments of Neurology and Rehabilitation and Medicine, University of Washington (UW), Seattle, (2) physical therapists and occupational therapists from the Department of Rehabilitation Medicine at the UW, and (3) members of the advisory board of UW's Aging Rehabilitation Research and Training Center funded by National Institute on Rehabilitation Research and Disability. Additional physical therapist panel members were recruited by posting an invitation on the American Physical Therapy Association Neurological Section Listserv.

Participants with MS were recruited through a website and print advertisements as well as from a disability registry maintained at the UW. Individuals in the registry who had MS were sent an invitation letter followed by a phone call to assess their interest in participation.

Procedures

Item pool reduction: We designed sorting and ranking procedures to quantify, from the perspective of clinicians and persons with MS, the relevance of items in the PROMIS fatigue bank. The full item bank consists of 95 items, and we judged this to be too many for either clinicians or individuals with MS to meaningfully rank. Study investigators reduced the pool of candidate items in a series of successive steps. We started by dropping 13 items that PROMIS adopted from a previously published measure that targeted fatigue in cancer [16]. The remaining 82 items have one of two response sets ("Never, Rarely, Sometimes, Often, Always" or "Not at all, A little bit, Somewhat, Quite a bit, Very much"). Many of these items differ little in content, varying only in response options. For example, one item asks, "To what degree did you have to limit your social activities because of your fatigue?" and provides response options of "Not at all, A little bit, Somewhat, Quite a bit, or Very much". Another item asks, "How often did your fatigue interfere with your social activities?" and provides response options of "Never, Rarely, Sometimes, Often, or Always". Our prior cognitive interviews with persons who had disabilities documented a preference for frequency-based response options [17]. Therefore, we dropped items with duplicate content favoring items with the response options, "Never, Rarely, Sometimes, Often, or Always".

Ratings by clinician experts: The resulting subset of 44 items (see "Appendix 1") was judged by investigators to be small enough in number to be reviewed and rated meaningfully by a sample of clinicians. The items were duplicated onto 44, 2 inch by 3.5 inch paper note cards. To simplify the rating procedure, clinicians were asked first to sort the items into two piles—one pile representing the more relevant items to fatigue in MS and another

representing the less relevant items. From the "more relevant" item pile, clinicians selected the three items they would choose if they could only ask their MS patients three questions about their fatigue. Clinicians recorded the item numbers corresponding to their selections. Next clinicians selected the item from those remaining that they would ask if they could ask just one more question. This procedure was repeated until they had chosen a total of 10 items. The three items selected as "most relevant" were assigned a ranking of "8". The fourth item selected was assigned a ranking of "7", and so on through the tenth item that was assigned a ranking of "1". Thus, higher values indicated stronger preference.

Ratings by persons with MS: To reduce response burden and the cognitive challenge of the task for persons with MS, a subset of 20 of the 44 items was identified by the study team for administration to participants with MS. The 20 items are identified in "Appendix 1" (marked with an 'x' in the second column). The items were chosen in an effort to maximize: (a) content coverage and (b) inclusion of items with higher clinician rankings. Participants with MS were mailed the items on paper note cards. MS participants then followed written instructions to identify and record the three items of the 20 that were the most relevant to their fatigue. Instructions stated, "If your doctor were to ask you three (only 3) questions about your fatigue, what 3 questions would give your doctor the best description of your fatigue?" Paralleling the procedure for clinicians, participants with MS were asked to add one item at a time until their top 10 were identified. The procedure used for scoring clinician ranks also was used in scoring the rankings of participants with MS.

Open-ended questions: After participants ranked their top 10 items they were asked, "Are there other questions (ones not printed on the cards) that you think are needed for a good summary of your fatigue?" Space was provided for responses, and these were used to evaluate and compare the content validity of the PROMIS-Fatigue$_{MS}$ and the PROMIS-Fatigue$_{SFv1.0}$ (described in Objective 2).

Analysis—The rankings provided by clinicians and participants with MS were ordinal-level, not interval-level, data; therefore, the appropriate average of these ranks is the median. However, because only 10 items per person received a rank (all others were scored as zero), the median clinician rankings for most items were zero. To better discriminate among item ranks, we calculated the arithmetic mean rank across raters, referred to hereafter as the "relevance index" (RI). Two RIs were calculated. One RI value was based on the rankings of the 20 items by participants with MS (RI-MS$_{20}$), and the second RI value was based on expert clinician rankings of the 44 items (RI-EX$_{44}$). We note that this index provides relative (not equal-interval-level) information about the strength of clinician and MS participant preferences for one item over another. Combining the results from the two ranking procedures, we selected items to constitute the PROMIS-Fatigue$_{MS}$. Items for the PROMIS-Fatigue$_{MS}$ were chosen to maximize three criteria: (1) high rankings by participants with MS, (2) high discrimination parameters based on IRT calibration, and (3) cumulatively representativeness of the content of the full PROMIS Fatigue item bank.

Objective 2: in a new sample of individual with MS, compare the known groups and content validity of scores on the PROMIS-Fatigue$_{MS}$ and the PROMIS-Fatigue$_{SFv1.0}$

Participants—Potential participants were recruited from among participants in an ongoing, longitudinal survey of individuals with MS. The survey was designed to assess quality of life of persons over time. Participants for the longitudinal study were recruited from the mailing list of the Western Washington chapter of the National MS Society. The first six administrations were at 4-month intervals; the remainder was administered every 7–11 months. A detailed description of the study sample has been published elsewhere [18]. For the current study, 242 of the 584 who completed the survey at the seventh time point were invited to participate.

Instrumentation—Participants with MS completed the items of the PROMIS-Fatigue$_{SFv1.0}$ and the items of the PROMIS-Fatigue$_{MS}$. In addition, they responded to the fatigue severity item, "To what degree have you experienced fatigue" [19]. The item was scored on a 0-–10 numerical rating scale where 0 = "not at all" and 10 = "a great deal".

All participants had completed demographic and symptom questionnaires within the past 3–9 months; these questionnaires were not repeated in the current study. In the analyses, we used additional data collected at a previous time point of the longitudinal survey, i.e., the self-report version of the Expanded Disability Status Scores (EDSS) [20] and a one-item measure of vitality taken from Medical Outcomes Study Short Form [21]. This item asks, "During the past 4 weeks, how much energy did you have? (Very much, quite a lot, some, a little, none)."

Analyses—Descriptive analyses (e.g., means, standard deviations) were completed on demographic and clinical variables. Summed scores on the PROMIS-Fatigue$_{SFv1.0}$ were converted to their PROMIS T-score equivalents based on a PROMIS-provided concordance table [http://www.assessmentcenter.net]. To score the new PROMIS-Fatigue$_{MS}$ SF, we derived a new concordance table that associates summed scores on the PROMIS-Fatigue$_{MS}$ with scores on the PROMIS T-score metric (Appendix 2). Score conversion was accomplished using the program IRTScore [22, #35] that converts raw scores to their IRT-calibrated, score equivalents (theta). To convert scores to the PROMIS-Fatigue$_{MS}$ T-score metric, theta values were multiplied by 10 and 50 was added. Association among scores on the two short forms was estimated by calculating Pearson correlation coefficient, and a Bland–Altman plot [23] was developed to compare score concordance.

Known groups validity: To evaluate and compare the validity of PROMIS-Fatigue$_{MS}$ scores and scores from the PROMIS-Fatigue$_{SFv1.0}$, we assessed "known groups" validity with respect to fatigue severity, vitality, and EDSS mobility score. We hypothesized that persons with higher fatigue scores on PROMIS-Fatigue$_{MS}$ and PROMIS-Fatigue$_{SFv1.0}$ should have higher reported fatigue severity, lower vitality scores, and higher EDSS scores (greater disability). These hypotheses were evaluated using analysis of variance (ANOVA) to compare means by group. For the comparison based on fatigue severity, the 0–10 ratings and published cutoff recommendations [24] were used to divide the sample into those with no fatigue (0), mild fatigue (1), moderate fatigue (2–4), and severe fatigue (5–10). The sample

also was divided into four groups based on answers to the SF8 vitality item, "During the past 4 weeks, how much energy did you have?" ("None or A Little, Some, Quite a Lot, Very Much"). Those responding "None" and "A Little" were combined for these analyses because only two participants endorsed the "None" category. EDSS scores were collapsed into three groups based on gross mobility levels (0–4.0 = mild, 4.5–6.5 = moderate, 7.0–9.5 = severe) [20].

Content validity: The requirements of content validity are met when the items of a measure cover the range of content relevant to the population of interest. Responses to the item, "what other information should be asked in addition to the top ten questions you selected" were independently categorized according to recurring themes by two of the authors (Cook and Bamer). Subsequently, the investigators met and came to consensus regarding the names and number of organizing categories. Next, the content of the PROMIS-Fatigue$_{MS}$ and the PROMIS-SF$_{V1.0}$ was compared to these categories to evaluate content coverage.

Results

Objective 1: identify items for the PROMIS-Fatigue$_{MS}$ relevant to the experience of fatigue in MS

Participants—A total of 37 clinicians completed the modified sorting exercise with the 44 PROMIS fatigue items—27 PTs, 7 MDs, and 3 Ots. Years of experience treating persons with MS ranged from 2 to 37 years (n = 36; mean = 15.8; SD = 8.9).

Of the 31 invited individuals with MS from the UW registry, 21 (68%) agreed to participate and completed the sorting procedure. Forty-one additional individuals with MS saw a study advertisement or heard about the study from someone else and contacted the researchers directly. Of these 41, 25 (61%) subsequently completed the sorting procedure, for a grand total of 46 participants with MS. Characteristics of the participants are reported in Table 1.

Finalizing the PROMIS-Fatigue$_{MS}$—RI values were calculated based on clinician ratings of 44 candidate items (RI-CL$_{44}$). The results are reported in "Appendix 1". The ten items with the highest clinician rankings encompassed most of the available content areas in the PROMIS Fatigue item bank. These ten items were included in the reduced candidate set presented to participants with MS. An additional 10 items were selected to maximize three selection criteria: (1) high rankings by clinicians, (2) high discriminations, and (3) cumulatively representativeness of the content of the PROMIS Fatigue item bank. The final 20 items chosen to be administered to participants with MS are indicated in "Appendix 1".

After the individuals with MS ranked the 20-item subset, their RI values were calculated and compared to those of clinicians (Appendix 1). There were both consistency and divergence in the RI values given by clinicians and by persons with MS. For example, both groups gave higher ratings to the item, "To what degree did your fatigue interfere with your physical functioning", and lower ratings to the item, "How often were you too tired to take a bath or shower?" In contrast, the item, "How often did you feel tired even when you hadn't done anything?" was given much higher ratings by persons with MS than by clinicians.

When selecting the final set of items for PROMIS-Fatigue$_{MS}$ short form, we relied on the ratings of persons with MS and included their top six rated items. In addition, we added the item, "How often did your fatigue interfere with your social activities?" (FATIMP4). This item was included to extend the content coverage of the short form. An eighth item also was added, "How often were you too tired to enjoy life? (FATEXP26)." This item was included after review of participants' open-ended responses (described below). This item targets the emotional impact of fatigue, an area many participants' indicated was "needed for a good summary" of their fatigue. The eight items of the PROMIS-Fatigue$_{MS}$ and the seven items included in the original PROMIS-Fatigue$_{SFv.1.0}$ are reported in Table 2. A concordance table that associates summed scores on the PROMIS-Fatigue$_{MS}$ with scores on the PROMIS T-score metric is provided as "Appendix 2".

Objective 2: in a new sample of individual with MS, compare the known groups and content validity of scores on the PROMIS-Fatigue$_{MS}$ and the PROMIS-Fatigue$_{SFv1.0}$

Participants—A total of 231 out of the 242 (95%) invited individuals completed the survey. Table 1 summarizes their demographics and clinical characteristics.

Analyses—Scores from the PROMIS-Fatigue$_{SFv1.0}$ and the PROMIS-Fatigue$_{MS}$ were highly correlated ($r = 0.92$). Figure 1 is a Bland–Altman plot [23] that shows the correspondence between scores. As the plot shows, 95% of the differences between the two short forms were within approximately 6 points of each other. Six points on the PROMIS T-score metric are 0.6 standard deviations (SDs) suggesting substantial concordance between scores on the two SFs.

<u>**Known groups validity:**</u> To evaluate and compare the known groups validity of scores from the PROMIS-Fatigue$_{SFv1.0}$ and the PROMIS-Fatigue$_{MS}$, we divided our sample into groups we expected to differ on self-reported fatigue. As described in the methods, the sample was categorized with respect to fatigue severity, vitality, and EDSS group. Mean fatigue scores were analyzed using ANOVA. On the 0–10 fatigue severity rating scale, only 6 individuals recorded a score of "0" (no fatigue), and only 12 recorded a score of "1" (mild fatigue). These two groups were collapsed into a single category for the known groups analyses. Similarly, on the Medical Outcomes Short Form vitality item, only two respondents chose the response "none", so the categories of "none" and "a little" were combined for the known groups analyses. For all comparisons, the omnibus hypothesis of no differences among groups was rejected with P-values < 0.0001 providing evidence for known-group validity. Table 3 presents the group-level means and standard deviations. All mean values were in the expected direction except for one. Mean PROMIS Fatigue$_{MS}$ scores for those who were moderate and severe on the EDSS were virtually the same (60.7 and 60.5, respectively). The scores of neither short form exhibited an advantage over the other in these analyses.

<u>**Content validity:**</u> Of the 46 participants with MS who ranked the 20 candidate fatigue items, 30 participants made a total of 67 item suggestions in response to the query, "What other information should be asked in addition to the top ten questions you selected." Complete results are available from the authors. Table 4 reports the categories and the number of comments per category. In summary, the category that drew the most responses

was "impact of fatigue on activities of daily living (ADLs) and instrumental ADLS (IADLs)". For example, one respondent suggested the item, "Because of your fatigue, do you find your personal hygiene sliding?" Eleven comments pertained to "fatigue triggers" (e.g., "How does the heat affect your fatigue.") Ten comments each pertained to "coping strategies" (e.g., "How often do you plan a nap into your day in order to have the energy to do an activity later in the day?"), "emotional impact" of fatigue (e.g., "How often does fatigue cause short temper?"), and "temporal aspects" of fatigue (e.g., "What time of day are you most fatigued?"). Five comments each related to "cognitive impact" of fatigue ("Do you feel at times you are in a mental fog?") and interaction between fatigue and other symptoms (e.g., "Does pain interfere with fatigue?"). One respondent suggested asking about impact of fatigue on sex and another commented, "I think there needs to be an opportunity to comment with greater detail."

Also included in Table 4 are the items from the generic and new MS-targeted short forms that target different categories of open-ended responses. The PROMIS-Fatigue$_{MS}$ had somewhat better content validity because of the addition of the item, "How often were you too tired to enjoy life?"–an item that targets the emotional impact of fatigue.

Six of the categories were not targeted by either short form. Much of the recommended content related to aspects of fatigue that would be important to share with a clinician in the context of an office visit (e.g., "Have you noticed any patterns/triggers for your fatigue?") or would be important in measuring latent constructs other than fatigue such as coping strategies. Inclusion of all these constructs would not be practical in a single, unidimensional measure. Though this content clearly was important to at least some respondents, we judged it not to be appropriate for inclusion in the measurement of our targeted construct.

There was one exception. One respondent suggested including an item about impact of fatigue on sex. Such items were included among the items tested for inclusion in the PROMIS fatigue bank, but such items tend to draw large numbers of missing responses, perhaps because these items are applicable to many but not all. We chose not to include an item about impact of fatigue on sex in the short form.

Discussion

Limitations

The current study has limitations. First, our measures of vitality and disability (EDSS) were collected 3–9 months prior to the current study. A stronger design would have been to re-administer these measures when collecting the data for the current study. Second, a cross-validation of the content validity of the PROMIS-Fatigue$_{MS}$ in a new sample of participants would have been useful to evaluate the generalizability of our findings.

A potential limitation of the PROMIS-Fatigue$_{MS}$ is our decision not to include an item about the impact of fatigue on sex, though this concern was raised by one of our participants. We recommend that, when appropriate, the impact of fatigue be addressed in clinical settings and that optional items addressing sex and fatigue be included in survey studies.

Conclusions

The state of the art in the development of self-reported outcomes measures has shifted from static instruments developed using classical test theory approaches to the development of item banks using IRT. Recent national initiatives such as PROMIS [25] have constructed item banks intended for a range of chronic conditions. However, it has been argued that the content validity of a measure must be demonstrated in every patient population in which it will be used [15]. In the current study, we developed the PROMIS-Fatigue$_{MS}$, a short form for measuring the fatigue of persons who have MS and ensured its content validity by using items that individuals with MS considered most relevant. The validity of the measure was further supported by the ability of PROMIS-Fatigue$_{MS}$ items to discriminate among groups expected to differ in levels of fatigue.

We recommend the use of this new short form when studying fatigue in individuals with MS. This recommendation, however, is not based on demonstrated superiority of the scores of the PROMIS-Fatigue$_{MS}$ compared to those of the PROMIS-Fatigue$_{SFv1.0}$. Mean scores on the two measures varied by less than one point in all known groups validity comparisons. However, our short form development strategy amounted to a "vetting" of specific PROMIS fatigue items by persons who have MS, easing concerns clinicians and researchers might have about the generalizability of PROMIS fatigue items in this population.

Researchers disagree about how critical it is to validate, in every clinical population, outcome measures that are common across many diseases. It is accepted that symptoms and outcomes such as fatigue, pain, and dysfunction may be precipitated by different disease processes in different clinical populations. Some argue, however, that there is more commonality than differences in patients' experiences of symptoms and outcomes and in patients' perceptions of their impact on quality of life. Others contend that content validity must be established in every population in which a measure will be used. PROMIS has taken the former position maintaining that measures can be developed whose scores are valid across a range of chronic conditions. In the current study, scores from the PROMIS-Fatigue$_{MS}$ and PROMIS-Fatigue$_{SFv1.0}$ discriminated equally well with respect to known groups validity, despite the fact that the former was developed for use across a range of conditions, and the latter was designed to be particularly relevant to persons with MS. Future testing of PROMIS instruments in different clinical populations and the aggregation of results will provide empirical evidence to inform this important debate.

Abbreviations

MS
Multiple sclerosis

PROMIS-Fatigue $_{SFv1.0}$
PROMIS 7-item fatigue short form, Version 1.0

PROMIS
Patient Reported Outcome Measurement Information System

U.S

United States

IRT

Item response theory

UW

University of Washington, Seattle

RI

Relevance index

RI-MS$_{20}$

RI value based on participant rankings of 20 PROMIS items

RI-EX$_{44}$

RI value based on expert clinician rankings of 44 PROMIS items

EDSS

Expanded Disability Status Scores

SF8

8-item Health Survey Short Form

References

1. National Multiple Sclerosis Society. Clinical study measures: Mental health inventory. National Multiple Sclerosis Society; 2008. [cited June 27, 2008]; Available from: http://www.nationalmssociety.org/for-professionals/researchers/clinical-study-measures/mhi/index.Aspx

2. Chwastiak LA, Gibbons LE, Ehde DM, Sullivan M, Bowen JD, Bombardier CH, et al. Fatigue and psychiatric illness in a large community sample of persons with multiple sclerosis. Journal of Psychosomatic Research. 2005; 59(5):291–298. [PubMed: 16253619]

3. Kraft GH, Freal JE, Coryell JK. Disability, disease duration, and rehabilitation service needs in multiple sclerosis: Patient perspectives. Archives of Physical Medicine and Rehabilitation. 1986; 67(3):164–168. [PubMed: 3954578]

4. Amato MP, Ponziani G, Rossi F, Liedl CL, Stefanile C, Rossi L. Quality of life in multiple sclerosis: The impact of depression, fatigue and disability. Multiple Sclerosis. 2001; 7(5):340–344. [PubMed: 11724451]

5. Johnson SL. The concept of fatigue in multiple sclerosis. The Journal of Neuroscience Nursing. 2008; 40(2):72–77. [PubMed: 18481736]

6. Fisk JD, Pontefract A, Ritvo PG, Archibald CJ, Murray TJ. The impact of fatigue on patients with multiple sclerosis. Canadian Journal of Neurological Sciences. 1994; 21(1):9–14. [PubMed: 8180914]

7. O'Connor AB, Schwid SR, Herrmann DN, Markman JD, Dworkin RH. Pain associated with multiple sclerosis: Systematic review and proposed classification. Pain. 2008; 137(1):96–111. [PubMed: 17928147]

8. Pompeii LA, Moon SD, McCrory DC. Measures of physical and cognitive function and work status among individuals with multiple sclerosis: A review of the literature. Journal of Occupational Rehabilitation. 2005; 15(1):69–84. [PubMed: 15794498]

9. Hubsky EP, Sears JH. Fatigue in multiple sclerosis: Guidelines for nursing care. Rehabilitation Nursing. 1992; 17(4):176–180. [PubMed: 1631394]

10. Cook KF, O'Malley KJ, Roddey TS. Dynamic assessment of health outcomes: Time to let the cat out of the bag? Health Services Research. 2005; 40(5 Pt 2):1694–1711. [PubMed: 16179003]

11. Cella D, Yount S, Rothrock N, Gershon R, Cook K, Reeve B, et al. The patient-reported outcomes measurement information system (PROMIS): Progress of an NIH roadmap cooperative group during its first two years. Medical Care. 2007; 45(5 Suppl 1):S3–S11.

12. Rothrock N, Hays RD, Spritzer K, Yount SE, Riley W, Cella D. Relative to the general us population, chronic diseases are associated with poorer health-related quality of life as measured by the patient-reported outcomes measurement information system (PROMIS). Journal of Clinical Epidemiology. 2010; 63(11):1195–1204. [PubMed: 20688471]

13. Riley WT, Rothrock N, Bruce B, Christodolou C, Cook K, Hahn EA, et al. Patient-reported outcomes measurement information system (PROMIS) domain names and definitions revisions: Further evaluation of content validity in irt-derived item banks. Quality of Life Research. 2010; 19(9):1311–1321. [PubMed: 20593306]

14. Lai J-S, Cella D, Choi S, Junghaenel DU, Christodoulou C, Gershon R, et al. How item banks and their applications can influence measurement practice in rehabilitation medicine: A PROMIS fatigue item bank example. Archives of Physical Medicine and Rehabilitation. (in press).

15. U.S. Department of Health and Human Services Food and Drug Administration. 2009. http://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm193282.pdf

16. Yellen SB, Cella DF, Webster K, Blendowski C, Kaplan E. Measuring fatigue and other anemia-related symptoms with the functional assessment of cancer therapy (FACT) measurement system. Journal of Pain and Symptom Management. 1997; 13(2):63–74. [PubMed: 9095563]

17. Amtmann D, Cook KF, Johnson KL, Cella D. The PROMIS initiative: Examples of applications in rehabilitation. Archives of Physical and Medical Rehabilitation. (in press).

18. Bamer AM, Johnson KL, Amtmann DA, Kraft GH. Beyond fatigue: Assessing variables associated with sleep problems and use of sleep medications in multiple sclerosis. Clinical Epidemiology. 2010; 2010(2):99–106. [PubMed: 20838467]

19. Belza BL, Henke CJ, Yelin EH, Epstein WV, Gilliss CL. Correlates of fatigue in older adults with rheumatoid arthritis. Nursing Research. 1993; 42(2):93–99. [PubMed: 8455994]

20. Bowen J, Gibbons L, Gianas A, Kraft GH. Self-administered expanded disability status scale with functional system scores correlates well with a physician-administered test. Multiple Sclerosis. 2001; 7(3):201–206. [PubMed: 11475445]

21. Ware, JE., Kosinski, M., Dewey, JE., Gandek, B. A manual for users of the sf-8 health survey. Lincoln, RI: Quality Metric Incorporated; 2001.

22. Flora, DB., Thissen, D. Electronic Research Memorandum #2002-1. Chapel Hill, NC: University of North Carolina, L.L. Thurstone Psychometric Laboratory; 2002. User's guide for IRTScore: Item response theory score approximation Software.

23. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986; 1(8476):307–310. [PubMed: 2868172]

24. Given B, Given CW, Sikorskii A, Jeon S, McCorkle R, Champion V, et al. Establishing mild, moderate, and severe scores for cancer-related symptoms: How consistent and clinically meaningful are interference-based severity cut-points? Journal of Pain and Symptom Management. 2008; 35(2):126–135. [PubMed: 18158231]

25. Cella D, Riley W, Stone A, Rothrock N, Reeve B, Yount S, et al. The patient-reported outcomes measurement information system (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. Journal of Clinical Epidemiology. 2010; 63(11):1179–1194. [PubMed: 20685078]

Appendix 1

See Table 5.

Appendix 2

See Table 6.
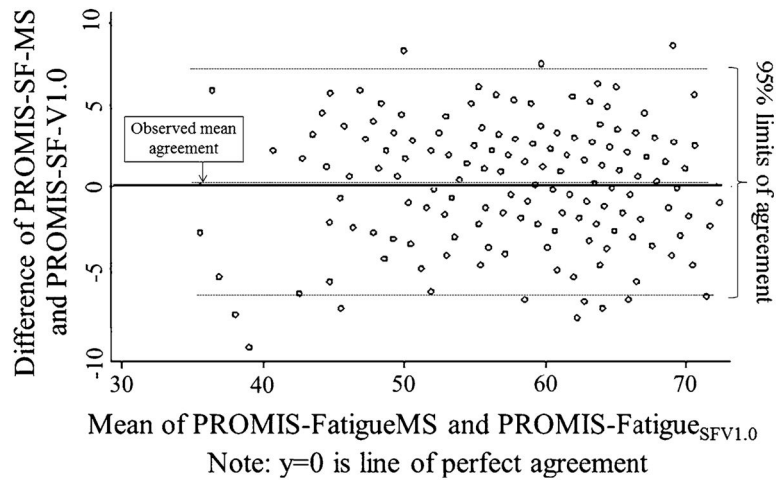
Fig. 1.
Bland-Altman plot showing the correspondence between scores on the PROMIS-Fatigue$_{SFV1.0}$ and PROMIS-Fatigue$_{MS}$

Table 1

Demographics of participants in sorting procedure (Objective 1) and validation study (Objective 2)

	Sorting procedure (Objective 1) Participants ($N = 46$)	Validation study (Objective 2) Participants ($N = 231$)
Age (years)	54.1 ± 9.3	53.6 ± 10.8
Disease duration (years)	14.1 ± 8.2	16.3 ± 9.5
Female	37 (80.4)	191 (82.7)
Fatigue (0–10 scale)		
None (0)	2 (4.3)	6 (2.6)
Mild(1)	0 (0)	12 (5.2)
Moderate (2–4)	5 (10.8)	58 (25.1)
Severe (5–10)	39 (84.8)	155 (67.1)
Disease severity (EDSS group)		
Mild (0–4.0)	N/A	83 (35.9)
Moderate (4.5–7.5)	N/A	104 (45.0)
Severe (8.0–9.5)	N/A	44 (19.1)

Table 2

Items of the PROMIS-Fatigue$_{MS}$ and the PROMIS-Fatigue$_{SFv1.0}$

PROMIS-Fatigue$_{MS}$		PROMIS-Fatigue$_{SFv1.0}$	
PROMIS variable name	Item content	PROMIS variable name	Item content
FATIMP30	How often were you too tired to think clearly?	FATIMP30	How often were you too tired to think clearly?
FATEXP48	How often did you find yourself getting tired easily?	FATEXP18	How often did you run out of energy?
FATEXP6	How often did you feel tired even when you had not done anything?	FATEXP20	How often did you feel tired?
FATIMP16	How often did you have trouble finishing things because of your fatigue?	FATEXP5	How often did you experience extreme exhaustion?
FATIMP3	How often did you have to push yourself to get things done because of your fatigue?	FATIMP21	How often were you too tired to take a bath or shower?
FATIMP4	How often did your fatigue interfere with your social activities?	FATIMP40	How often did you have enough energy to exercise strenuously?
FATIMP49	To what degree did your fatigue interfere with your physical functioning?	FATIMP33	How often did your fatigue limit you at work (include work at home)?
FATEXP26	How often were you too tired to enjoy life?		

Table 3

Mean PROMIS Fatigue$_{SFv1.0}$ and PROMIS Fatigue$_{MS}$ scores by disability status, fatigue severity, and vitality scores

	N	PROMIS Fatigue$_{SF v1.0}$		PROMIS Fatigue$_{MS}$	
		Mean	SD	Mean	SD
Expanded disability status scale (EDSS)					
Mild (0–4)	83	52.2	8.2	52.5	9.2
Moderate (4.5–6.5)	104	60.5	6.4	60.7	5.6
Severe (7.0–9.5)	43	60.7	8.3	60.5	8.7
Fatigue severity (0–10 numerical rating scale)					
None/Mild (0–1)	18	43.0	4.5	42.5	5.4
Moderate (2–4)	58	51.0	6.0	51.3	6.6
Severe (5–10)	154	61.7	5.8	61.9	5.5
Vitality (item from the medical outcomes survey)					
None/A little	52	63.8	5.3	64.2	5.4
Some	88	59.9	6.3	60.1	5.5
Quite a lot	44	55.7	6.6	56.0	6.8
Very Much	45	47.5	7.3	47.0	7.9
Missing	1				

Table 4

Participant comments compared to short forms

Category	Number of responses	PROMIS-Fatigue$_{SFv1.0}$	PROMIS-Fatigue$_{MS}$
*ADL/IADL Impact	14	How often did your fatigue limit you at work (include work at home)?	How often did your fatigue limit you at work (include work at home)?
		How often were you too tired to take a bath or shower?	How often did you have trouble finishing things because of your fatigue?
Cognitive	5	How often were you too tired to think clearly?	How often were you too tired to think clearly?
Emotional impact	10	Not addressed	How often were you too tired to enjoy life?
Coping	10	Not addressed	Not addressed
Temporal aspects	10	Not addressed	Not addressed
Symptom interaction	5	Not addressed	Not addressed
Triggers	11	Not addressed	Not addressed
General comment	1	Not addressed	Not addressed
Impact on sex	1	Not addressed	Not addressed

ADL activities of daily living, *IADL* instrumental activities of daily living

Table 5

Fatigue item variable name, content, relevance indices, and short form membership (denoted with an 'x')

PROMIS variable name	Item content	RI-CL$_{44}$	RI-MS$_{20}$	PROMIS-Fatigue$_{SFv1.0}$	PROMIS-Fatigue$_{MS}$
FATEXP6	How often did you feel tired even when you had not done anything?	0.92	4.39		X
FATIMP3	How often did you have to push yourself to get things done because of your fatigue?	1.78	4.37		X
FATIMP16	How often did you have trouble finishing things because of your fatigue?	1.89	3.82		X
FATIMP49	To what degree did your fatigue interfere with your physical functioning?	4.05	3.67		X
FATEXP48	How often did you find yourself getting tired easily?	1.30	3.37		X
FATEXP30	How often were you too tired to think clearly?	2.16	3.26	X	X
FATIMP33	How often did your fatigue limit you at work (include work at home)?	3.95	3.17	X	
FATEXP7	How often did you feel your fatigue was beyond your control?	1.00	2.76		
FATIMP14	How often did your fatigue make it difficult to organize your thoughts when doing things at work (include work at home)?	1.41	2.63		
FATEXP26	How often were you too tired to enjoy life?	1.70	2.43		X
FATIMP17	How often did your fatigue make it difficult to make decisions?	1.32	2.13		
FATEXP34	How tired did you feel on average?	1.73	2.09		
FATIMP9	How often did you fatigue make it difficult to plan activities ahead of time?	1.62	2.09		
FATEXP21	How fatigued were you when your fatigue was at its worst?	1.84	2.07		
FATIMP4	How often did your fatigue interfere with your social activities?	2.68	1.93		X
FATEXP5	How often did you experience extreme exhaustion?	1.19	1.93	X	
FATIMP29	How often were you too tired to leave the house?	1.38	1.91		
FATIMP21	How often were you too tired to take a bath or shower?	1.81	1.34	X	
FATIMP15	How often did your fatigue interfere with your ability to engage in recreational activities?	1.05	1.15		
FATIMP26	How often were you too tired to socialize with your family?	1.24	0.67		
FATIMP6	How often did your fatigue make you feel slowed down in your thinking?	1.43			
FATEXP22	How often were you bothered by your fatigue?	1.27			
FATIMP19	How often were you too tired to do your household chores?	1.27			
FATIMP56	How often were you too tired to socialize with your friends?	1.08			
FATIMP24	How often did you have trouble starting things because of your fatigue?	0.86			
FATIMP25	How often was it an effort to carry on a conversation because of your fatigue?	0.81			
FATEXP20	How often did you feel tired?	0.78		X	
FATEXP49	How often did you think about your fatigue?	0.76			

PROMIS variable name	Item content	RI-CL$_{44}$	RI-MS$_{20}$	PROMIS-Fatigue$_{SFv1.0}$	PROMIS-Fatigue$_{MS}$
FATEXP51	How easily did you find yourself getting tired on average?	0.76			
FATIMP53	How often were you too tired to take a short walk?	0.76			
FATIMP42	How often were you less effective at home due to your fatigue?	0.73			
FATIMP55	How often did you have to force yourself to get up and do things because of your fatigue?	0.70			
FATIMP13	How often were you too tired to do errands?	0.65			
FATEXP50	How fatigued were you on the day you felt least fatigued?	0.59			
FATEXP19	How often were you physically drained?	0.54			
FATEXP54	How often did you have physical energy?	0.54			
FATIMP20	How often did your fatigue make you feel less alert?	0.43			
FATIMP10	How often did your fatigue make it difficult to start anything new?	0.38			
FATEXP18	How often did you run out of energy?	0.32	X		
FATIMP11	How often did your fatigue make you more forgetful?	0.30			
FATEXP42	How much mental energy did you have on average?	0.27			
FATEXP46	On how many days was your fatigue worse in the morning?	0.27			
FATIMP40	How often did you have enough energy to exercise strenuously?	0.24	X		
FATIMP8	How often were you too tired to watch television?	0.22			

Relevance indices were calculated as the arithmetic mean rank across raters (RI-CL$_{44}$ for clinician raters, RI-MS$_{20}$ for raters with MS)

Table 6

PROMIS-Fatigue_{MS} Raw score (Sum of All Item Scores) and their equivalent T-scores

PROMIS-Fatigue_{MS} raw score	Equivalent PROMIS T-score
8	34.1
9	39.3
10	41.8
11	43.6
12	45.1
13	46.4
14	47.6
15	48.7
16	49.8
17	50.9
18	52.0
19	53.0
20	54.1
21	55.1
22	56.2
23	57.3
24	58.3
25	59.3
26	60.4
27	61.5
28	62.5
29	63.6
30	64.7
31	65.8
32	66.9
33	68.1
34	69.3
35	70.5
36	71.9
37	73.4
38	75.1
39	77.5
40	80.9