Published in final edited form as:

Mult Scler. 2014 July ; 20(8): 1102–1111. doi:10.1177/1352458513517279.

Reducing survey burden: Feasibility and validity of PROMIS measures in multiple sclerosis

Angela Senders^{1,2}, Douglas Hanes², Dennis Bourdette^{1,3}, Ruth Whitham^{1,3}, and Lynne Shinto¹

¹Department of Neurology, Oregon Health & Science University, Portland, OR 97239, USA

²Helfgott Research Institute, National College of Natural Medicine, Portland, OR 97201, USA

³Department of Neurology, Portland Veterans Affairs Medical Center, Portland, OR 97239, USA

Abstract

Background—Patient-reported outcomes are important for clinical research and care, yet administering and scoring questionnaires requires considerable effort and time. The Patient Reported Outcomes Measurement Information System (PROMIS) could considerably reduce administrative obstacles and lessen survey burden for participants.

Objective—Assess feasibility and validity of PROMIS compared to commonly used legacy measures for multiple sclerosis (MS).

Methods—Cross-sectional survey; 133 participants with confirmed MS completed legacy surveys and PROMIS Computerized Adaptive Tests (CATs) for depression, anxiety, pain, fatigue, and physical function. A multitrait-multimethod analysis was conducted and verified with confirmatory factor analysis.

Results—Correlations between PROMIS and corresponding legacy measures were large (0.67 to 0.87). Multitrait-multimethod criteria were generally well met, providing good evidence of validity of PROMIS measures. PROMIS surveys asked fewer questions and required substantially less time to complete than legacy scales.

Conclusions—Results provide evidence of construct validity of PROMIS for use with MS. Several aspects of PROMIS CATs make them an important resource, including: 1) less time required to complete, 2) reduced missing data, and 3) automatic scoring referenced to the general population. Findings support the use of PROMIS in MS research and may have broader implications for clinical care, as well.

Keywords

multiple sclerosis; outcome measurement; self report; PROMIS; construct validity

Corresponding author Angela Senders, Department of Neurology, Oregon Health & Science University, 3181 SW Sam Jackson Park Road, Mail Code CR120, Portland, Oregon, 97239. Phone: 503-494-7798 Fax: 503-494-0966 senders@ohsu.edu.

Conflict of Interest Statement The authors declare that there is no conflict of interest.

Introduction

Patient-reported outcomes (PROs) quantify a person's experience with their health and the effect of treatment on how they feel or function. As the patient is the only reliable source of this information, PROs are important outcomes in both clinical care and research settings. There are, however, substantial obstacles to collecting PROs. Administering questionnaires, dealing with missing data, scoring surveys, and effectively storing data require considerable time from providers and researchers. Longer surveys can lead to respondent fatigue and place additional burden on patients with disease, potentially resulting in mistakes and lower completion rates (1). The Patient Reported Outcomes Measurement Information System (PROMIS) is a free, NIH funded, online repository of PROs with the novel capability of administering computerized adaptive tests (CATs) (2). CATs reduce survey length by selecting the most informative question based on the participant's prior response. In reducing the number of questions asked, PROMIS may improve the survey experience for participants, minimize the risk of missing data, and substantially reduce administration burden.

Multiple sclerosis (MS) research has long used legacy scales to collect PROs. In order to validate PROMIS CAT measures for future use in MS, we conducted a cross sectional survey (n=133) in which respondents completed both PROMIS and related legacy measures. Our objective was to explore the feasibility and validity of PROMIS compared to legacy measures for depression, anxiety, fatigue, pain, and physical function in MS. Results may not only be applicable to MS research but have broader implications for clinical care, as well.

Materials and Methods

The current analysis was part of a study whose primary aim was to assess the relationship between self-reported mindfulness and perceived stress in MS. The evaluation of PROMIS and legacy scales was established a priori as an exploratory aim. Sample size (n=150) was calculated to find correlations between primary outcomes of at least .30 with 80% power, an alpha = .05, and a 5% loss of data.

Design, Setting, and Study Participants

Recruitment began after approval from the Oregon Health & Science University (OHSU) Institutional Review Board. Participants were scheduled for one study visit to complete legacy and PROMIS questionnaires. A convenience sample of men and women were recruited during their outpatient visits to the MS Center at OHSU. The study was also advertised with the National MS Society local chapter and through MS community events. Inclusion criteria comprised any type of MS, ability to read and write in English, age 18 to 90 years old. Exclusion criteria included a relapse or exacerbation within the previous 90 days. MS diagnosis was confirmed by chart review according to 2010 McDonald criteria (3). Between December 2011 and February 2013, 150 people with MS gave written informed consent and completed the study.

Outcome Measures

All legacy questionnaires are validated surveys used in MS clinical trials. All outcome measures ask participants to rate their experience on a Likert-type scale. Legacy measures were administered by pen and paper; PROMIS measures were administered using a laptop computer.

Beck Depression Inventory I (BDI) is a 21-item survey. Total BDI scores range from 0-63; higher scores indicate higher levels of depression (4).

State-Trait Anxiety Inventory (STAI) consists of two 20-item subscales, one that assesses anxiety in the present moment (state) and one that assesses dispositional anxiety (trait) (5). For this analysis, only trait anxiety was used. The STAI trait subscale score ranges from 20-80; higher scores reflect more anxiety.

Modified Fatigue Impact Scale (MFIS) is a 21-item survey derived from the Fatigue Impact Scale (6,7). The total MFIS score ranges from 0 – 84; higher scores reflect more fatigue.

Medical Outcomes Study Pain Effects Scale (PES) is a 6-item scale assessing the effects of pain on behavior and mood over the past 4 weeks (7,8). Scores range from 6 – 30; higher scores indicate more pain.

Physical Function subscale of the Short Form – 36 (SF-36)—The SF-36 is a generic measure of health-related quality of life comprised of eight subscales (9,10). Subscales are scored on a T-scale metric with the mean score of a general U.S. population equal to 50 and a standard deviation of 10. Higher scores indicate better physical function.

PROMIS CAT 1.0 inventories included depression, anxiety, fatigue, pain interference, and physical function (11–13). Surveys are administered and scored via a web-based application hosted by the Assessment CenterSM. CATs were configured to administer enough items to achieve a standard error (SE) of <0.30 after a minimum of 4 questions were asked. Each CAT stopped after a maximum of 12 questions, regardless of whether the SE criterion was met. Scores are reported on a T-score metric normalized so that population mean equals 50 and standard deviation 10.

Statistical Analysis

Construct validity of PROMIS measures was evaluated using a multitrait-multimethod matrix (MTMM) (14). MTMM analysis involves correlating scores of different traits (e.g. depression, anxiety, pain, etc.) across different methods of investigation (e.g. PROMIS vs. legacy instruments). Strong correlations between two different methods measuring a single trait (e.g. PROMIS and legacy measures for depression) support convergent construct validity, whereas weaker correlations among measures of different traits are evidence of discriminant validity. Detailed MTMM criteria that must be fulfilled to support construct validity are described by Campbell and Fiske (14).

Preliminary analyses were performed using STATA 12. Descriptive analyses were used to summarize demographic information and outcomes for PROMIS and legacy instruments.

Cronbach's alpha was generated for all legacy surveys as a measure of internal consistency (15). Due to the CAT format, alpha for PROMIS measures cannot be computed from study data. Alphas for PROMIS scales applied to a more general population are provided by Assessment CenterSM, and are provided here for reference. An alpha > 0.80 is generally considered satisfactory as a measure of survey reliability (16). The following were used for the strength of correlation coefficients: 0.0 – 0.29 small, 0.30 – 0.49 moderate, 0.50 – 0.74 good, and $r > 0.75$ a strong relationship (17).

Comparison of correlation coefficients was considered successful if the validity coefficient was significantly larger than the comparison coefficient ($p < 0.05$) (18). Calculations of significance were programmed with R software v.2.15.1. Each comparison of correlations was tested using the maximum amount of data available ($n=133$). Due to an oversight in administering legacy surveys, Fatigue data is not available for 19 of these 133 participants. Thus, comparisons in which at least one correlation involved legacy Fatigue were made with data for 114 participants. To substantiate the MTMM analysis, confirmatory factor analysis (CFA) was conducted using R's sem package (19) to test for underlying trait and method effects. The principal model loaded PROMIS and Legacy measures onto corresponding latent trait factors of Depression, Anxiety, Fatigue, Pain, and Physical Function. For CFA, the 19 missing legacy Fatigue measures were imputed as the mean Fatigue scores of the other 114 participants. CFA was then conducted on data for all 133 participants who completed the PROMIS scales.

Results

Participant Characteristics

Two hundred fourteen people with MS were approached to participate in this study (Figure 1). Missing and excluded data resulted in sample sizes for individual scales that varied between 120 and 146. Demographic data is displayed in Table 1.

Descriptive statistics for legacy and PROMIS questionnaires are presented in Table 2. Surveys were administered in the same order for each visit. Because PROMIS was an exploratory aim, these questionnaires were administered at the end of the visit. Several participants experienced survey fatigue and were unable to complete the entire study ($n=13$); this left 133 participants who completed PROMIS and legacy surveys (Table 1). Non-completers reported greater levels of disability; 69% reported needing at least some kind of regular walking support compared to 26% of those that completed PROMIS. Additionally, there were more participants with primary or secondary progressive MS in the non-completer group (62% vs. 25%, Table 1). Legacy scores with available normative data were converted to a T score metric to enhance comparability with PROMIS data (Table 3).

Feasibility

Legacy questionnaires were completed by pen and paper; PROMIS questionnaires were completed online. Six participants required modifications to study administration due to poor vision or hand tremors: questions were read aloud to one participant who provided verbal responses that were recorded by study staff. Five other participants read each question

to themselves and verbally spoke their answers to study staff. There was no difference in modifications made for legacy and PROMIS scales.

PROMIS asked fewer questions compared to corresponding legacy instruments, with the exception of the pain surveys (Table 2). The average time to complete all five PROMIS surveys was 4.08 minutes. Participants were not timed during legacy surveys, although estimated time to complete all five legacy surveys was 15.6 minutes (78 questions estimated at five questions per minute) (1).

Multitrait-Multimethod Analysis

The MTMM matrix is provided in Table 4. Reliability coefficients (*italics*) are consistently strong (ranging from .91 to .97) and, as expected, are the largest coefficients in the matrix. Validity coefficients (**bolded**) are good to strong (ranging from .67 to .87). Validity coefficients should be higher than correlations in their row and column within the PROMIS vs. Legacy quadrant of the matrix. Fatigue, pain, and physical function met this criterion 100% of the time; depression 87.5%; and anxiety 75% of the time. Validity coefficients should be higher than all off-diagonal coefficients in the PROMIS vs. PROMIS and Legacy vs. Legacy quadrants. Pain and physical function met this criterion 100% of the time; fatigue 95%; depression 85%; and anxiety 60% of the time. Significant interrelationships among traits are present within the matrix. This is acceptable as the pattern of trait interrelationship appears to be consistent between methods (14).

Confirmatory Factor Analysis

To verify results of the MTMM analysis, confirmatory factor analysis determined whether covariance of the ten measured scales could be explained by covariance of the five latent factors of depression, anxiety, fatigue, pain, and physical function. The initial model loaded each legacy and PROMIS measure onto one of five corresponding latent trait factors. (E.g., the PROMIS Depression and Legacy Depression measures were allowed to load onto a single latent Depression factor.) This returned a chi-squared statistic of $\chi^2(25)=43.22$, $p=0.013$; a Goodness-of-fit index of 0.940; and RMSEA index of 0.074, indicating good fit to the observed MTMM matrix. All loadings onto trait factors were significant ($p < 10^{-25}$) and greater than 0.80, indicating strong convergent validity of PROMIS and Legacy measures. Factor loadings and correlations are available in the supplementary materials.

Additional CFA modeling with two method factors and no traits yielded a method-method correlation > 0.95 . This suggests that any variation seen in scores between PROMIS and legacy scales is not due to systematic differences between the two methods of administration.

PROMIS outcomes by MS subtype

Mean PROMIS scores categorized by MS subtype are presented in Figure 2. As expected, participants with progressive disease report lower levels of physical function than those with relapsing remitting MS. Correspondingly, Figure 3 displays physical outcomes (fatigue, pain, and physical function) categorized by self-report level of disability. Physical function scores progressively decline across all categories from 'no disability' to 'unable to walk'.

Fatigue and pain increase across the first four categories, from ‘no disability’ to ‘some support needed’, and then decline over the last two categories. Because there were few progressive participants in this study, these data should be viewed as descriptive and future studies should confirm results with larger numbers of progressive participants.

Discussion

Patient-reported assessments of depression, anxiety, fatigue, pain, and physical function are essential in the care of people with MS and are important outcomes in clinical trials. MS research has long used legacy instruments to collect this information, and before adopting novel PROMIS methodology, the relationship of PROMIS CAT measures to traditional surveys needs to be understood. We found strong correlations between corresponding scales (0.67 – 0.87). In addition, PROMIS measures asked fewer questions, required substantially less time to complete, and had no missing data compared to legacy scales.

One of the aims of PROMIS is to improve precision while reducing survey burden for participants. In general, longer surveys result in lower completion rates; this is particularly true for those with more advanced disease who may have difficulty sustaining attention, holding a pen, or focusing vision for long periods of time. Indeed, we lost 13 of 146 participants over the course of our survey for these reasons. Non-completers reported higher levels of disability than completers, however it was not our experience that people with advanced disability had more difficulty completing PROMIS compared to legacy scales. On the contrary, because it requires less time and fewer questions to complete, we believe that using PROMIS increases the likelihood that self-report data is collected from people across the disability spectrum. In total, completion of PROMIS scales required an average of 24.3 items, compared to the 78 items of corresponding legacy scales. It is unsurprising that PROMIS requires fewer items, given that each CAT was programmed to ask no more than 12 questions. Here the interesting result is *how many* fewer questions are required by PROMIS surveys. Even with a stopping rule of 12, the average number of questions asked per PROMIS survey was only 4.86. Our data suggest that PROMIS and legacy scales provide similar results, even though PROMIS surveys tend to be substantially shorter.

Reducing the number of questions asked while preserving reliability will improve the survey experience for participants and minimize the risk of missing data. The average completion time for five PROMIS scales was 4.08 minutes. As a rule of thumb, a typical response time for Likert-type questionnaires is three to five questions per minute (1). For five legacy scales with a total of 78 items, an estimated completion time would be 15-26 minutes. This rule of thumb applies to a general population, and response times could be substantially longer for those with cognitive or physical limitations. Thus, reducing survey length has great implications for clinical research in neurology.

Additional benefits of using PROMIS in this study included no missing data. Participants are automatically reminded to answer any question unintentionally left blank, and if the respondent deliberately skips a question the CAT will substitute alternative, equivalent questions. Although study staff reviewed each legacy scale upon completion, it was impossible to avoid missing data for pen and paper surveys. In addition, both raw and T

scores for CATs are immediately available from Assessment CenterSM, substantially reducing the time devoted to scoring and data entry. Under real-world conditions, this will inevitably lead to more reliable outcomes. Because surveys are free, quick to administer, and automatically scored while the patient is potentially still in the office, PROMIS CATs may also work well in a clinical care setting. One academic center has reported using PROMIS successfully without disrupting the flow of clinic, although more work is needed in this area (21).

While there are several advantages to using PROMIS, there are also limitations that should be considered. CATs require some kind of computer interface with real time internet access to upload responses to the Assessment CenterSM database. This may pose a logistical problem for some trials or clinical assessments. PROMIS surveys can be administered as static questionnaires and delivered by pen and paper format, but static forms must be hand graded, even when administered electronically. Finally, a significant challenge for utilizing PROMIS assessments in clinical care remains how to seamlessly integrate data into electronic health records without a substantial increase in clinical staff workload.

Correlations between corresponding PROMIS and legacy measures are strong, but not perfect. To the extent that our common understanding of health is often guided by legacy scales, the PROMIS measures may therefore convey somewhat less information. However, the MTMM analysis can only assess the amount of overlap between PROMIS and legacy scores, and not how well either method measures the true underlying characteristic of interest. It is unclear if any information is lost with PROMIS, and it is possible that PROMIS measures are actually more informative about the real state of the respondent. For example, others have demonstrated that PROMIS physical function questions better differentiate among people with higher physical function than the SF-36 subscale (22). Indeed, we found a discrepancy between the means of the PROMIS and legacy physical function scales (Table 3) that likely results from the PROMIS scale's lack of an artificial upper bound on rating of physical function. In this case, a less than perfect correlation between PROMIS and legacy physical function may indicate that PROMIS provides more accurate information than the legacy measure. Future validity studies may aim to assess differences in information capture between PROMIS and legacy scales.

Despite potential limitations, we found PROMIS feasible for use in clinical research with good evidence of construct validity. To establish validity for longitudinal studies, determining minimally important differences, test-retest reliability, and responsiveness to change in MS is necessary. Yost et al. suggest that minimally important differences for cancer participants using the static PROMIS scales depression, anxiety, fatigue, pain interference, and physical functioning range between 3 and 6 points (23). Broderick et al. have shown that 7-day test-retest reliability was strong ($r = .80$) for 98 osteoarthritis participants completing pain intensity, pain interference, physical functioning, and fatigue PROMIS scales. Finally, Fries et al. provide evidence of responsiveness to change for static PROMIS physical function forms over a 12-month period for people with rheumatoid arthritis (24). Based on these findings PROMIS has good potential for use in longitudinal studies, although MS specific psychometrics should continue to be established.

Few studies have evaluated the use of PROMIS in MS (25–27) and to our knowledge, this is the first study to examine the association between legacy and PROMIS CAT measures of depression, anxiety, fatigue, pain, and physical function in MS. Our study has limitations as most participants came from a single center and convenience sampling may have led to under- or over-representation of certain factors within our sample. Indeed, only 10.5% of our participants (14/133) reported moderate to severe depression (> 21 on the BDI). The point prevalence for depression in MS clinical samples is often estimated closer to 30% (28–30), potentially limiting the generalizability of our results.

We also did not time participants as they completed legacy scales and instead used a general rule of thumb to compare completion times. Nevertheless, our estimates were conservative for a general population and may have underestimated the true impact that PROMIS surveys could have in patients with neurological conditions. Finally, this was a cross-sectional study. Future longitudinal studies that collect PROMIS outcomes will help identify minimally important differences, responsiveness to change, and test-retest reliability in the MS population. Despite potential limitations, we found PROMIS feasible for use in MS research with good evidence of construct validity. Several benefits of PROMIS CATs outweighed the use of legacy scales, including: 1) fewer questions asked and less time required of respondents, 2) no missing data, 3) prompt automatic scoring referenced to the general population, and 4) comparability across studies and conditions. Results may not only be applicable to MS, but may also have broader implications for a range of neurological conditions in which survey burden is a major concern. In addition to being a research tool, clinicians also have easy access to PROMIS and may find benefit using these surveys in a clinical setting.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Study Funding This work was supported by the Oregon Clinical and Translational Research Institute [Grant number UL1 RR024140 from the National Center for Research Resources and the National Center for Advancing Translational Sciences of the National Institutes of Health]; the National Center for Complementary and Alternative Medicine of the National Institutes of Health [Grant numbers AT002688 and 2R25AT002878-05A1]; the Agency for Healthcare Research Quality [Grant number 5T32HS017582-05], and the Medical Research Foundation of Oregon. The content of this paper is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

1. Hays, RD.; Reeve, BB. International Encyclopedia of Public Health. Acad; San Diego: 2008. Measurement and Modeling of Health-Related Quality of Life; p. 241-52.
2. 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. J Clin Epidemiol. Nov; 2010 63(11):1179–94. [PubMed: 20685078]
3. Polman CH, Reingold SC, Banwell B, Clanet M, Cohen JA, Filippi M, et al. Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. Ann Neurol. Feb; 2011 69(2):292–302. [PubMed: 21387374]

4. Beck A, Ward C, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. *Arch Gen Psychiatry*. Jun.1961 4:561–71. [PubMed: 13688369]
5. Spielberger, C. Manual for the state-trait anxiety inventory STAI (form Y). Mind Garden; Palo Alto, CA: 1983.
6. Fisk JD, Ritvo PG, Ross L, Haase DA, Marrie TJ, Schlech WF. Measuring the functional impact of fatigue: initial validation of the fatigue impact scale. *Clin Infect Dis*. Jan; 1994 18(Suppl 1):S79–83. [PubMed: 8148458]
7. Fischer JS, LaRocca NG, Miller DM, Ritvo PG, Andrews H, Paty D. Recent developments in the assessment of quality of life in multiple sclerosis (MS). *Mult Scler*. Aug; 1999 5(4):251–9. [PubMed: 10467384]
8. Stewart, AL., editor. *Measuring Functioning and Well-Being: The Medical Outcomes Study Approach*. 1st ed. Duke University Press Books; 1991.
9. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care*. Jun; 1992 30(6):473–83. [PubMed: 1593914]
10. Brazier JE, Harper R, Jones NM, O’Cathain A, Thomas KJ, Usherwood T, et al. Validating the SF-36 health survey questionnaire: new outcome measure for primary care. *BMJ*. Jul 18; 1992 305(6846):160–4. [PubMed: 1285753]
11. Rothrock NE, 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). *J Clin Epidemiol*. Nov; 2010 63(11):1195–204. [PubMed: 20688471]
12. DeWalt DA, Rothrock N, Yount S, Stone AA. Evaluation of item candidates: the PROMIS qualitative item review. *Med Care*. May; 2007 45(5 Suppl 1):S12–21. [PubMed: 17443114]
13. Liu H, Cella D, Gershon R, Shen J, Morales LS, Riley W, et al. Representativeness of the Patient-Reported Outcomes Measurement Information System Internet panel. *J Clin Epidemiol*. Nov; 2010 63(11):1169–78. [PubMed: 20688473]
14. Campbell D, Fiske D. Convergent and discriminant validation by the multitrait multimethod matrix. *Psychol Bull*. Mar; 1959 56(2):81–105. [PubMed: 13634291]
15. Cronbach L. Coefficient alpha and the internal structure of tests. *Psychometrika*. 1951; 16(3):297–334.
16. Nunnally, J.; Bernstein, I. *Psychometric Theory*. 3rd ed. McGraw-Hill; New York, NY: 1994.
17. Cohen, J.; Cohen, P.; West, SG.; Aiken, LS. *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. Third. Routledge Academic; 2002.
18. Hayashi T, Hays RD. A microcomputer program for analyzing multitrait-multimethod matrices. *Behavior Research Methods*. 1987; 19(3):345–8.
19. Fox, J.; Nie, Z.; Byrnes, J. [cited 2013 May 20] *Structural Equation Models (sem)* [Internet]. Available from: <http://socserv.socsci.mcmaster.ca/jfox/>
20. Shinto L, Yadav V, Morris C, Lapidus JA, Senders A, Bourdette D. Demographic and health-related factors associated with complementary and alternative medicine (CAM) use in multiple sclerosis. *Mult Scler*. Feb; 2006 12(1):94–100. [PubMed: 16459725]
21. Khanna D, Maranian P, Rothrock N, Cella D, Gershon R, Khanna PP, et al. Feasibility and construct validity of PROMIS and “legacy” instruments in an academic scleroderma clinic. *Value Health*. Jan; 2012 15(1):128–34. [PubMed: 22264980]
22. Hays RD, Liu H, Spritzer K, Cella D. Item response theory analyses of physical functioning items in the medical outcomes study. *Med Care*. May; 2007 45(5 Suppl 1):S32–38. [PubMed: 17443117]
23. Yost KJ, Eton DT, Garcia SF, Cella D. Minimally important differences were estimated for six Patient-Reported Outcomes Measurement Information System-Cancer scales in advanced-stage cancer patients. *Journal of Clinical Epidemiology*. May; 2011 64(5):507–16. [PubMed: 21447427]
24. Fries J, Rose M, Krishnan E. The PROMIS of better outcome assessment: responsiveness, floor and ceiling effects, and Internet administration. *J Rheumatol*. Aug; 2011 38(8):1759–64. [PubMed: 21807798]
25. Becker H, Stuifbergen A, Morrison J. Promising New Approaches to Assess Cognitive Functioning in People with Multiple Sclerosis. *International Journal of MS Care*. 2012; 14(2):71–6. [PubMed: 23550178]

26. Askew RL, Kim J, Chung H, Cook KF, Johnson KL, Amtmann D. Development of a crosswalk for pain interference measured by the BPI and PROMIS pain interference short form. *Qual Life Res.* Mar 29.2013
27. Bamer A, Cook K, Amtmann D, Amtmann D. Reliability and validity of pain, fatigue, depression, and social roles participation PRO measures in populations of individuals aging with a physical disability. *Quality of life research.* 2012; 20(Suppl 1):66.
28. Patten SB, Berzins S, Metz LM. Challenges in screening for depression in multiple sclerosis. *Mult Scler.* Nov; 2010 16(11):1406–11. [PubMed: 20685765]
29. McGuigan C, Hutchinson M. Unrecognised symptoms of depression in a community-based population with multiple sclerosis. *J Neurol.* Feb; 2006 253(2):219–23. [PubMed: 16177840]
30. Chwastiak L, Ehde DM, Gibbons LE, Sullivan M, Bowen JD, Kraft GH. Depressive symptoms and severity of illness in multiple sclerosis: epidemiologic study of a large community sample. *Am J Psychiatry.* Nov; 2002 159(11):1862–8. [PubMed: 12411220]

Figure 1.
Study flow diagram.

Figure 2.

Mean PROMIS T scores for depression, anxiety, fatigue, pain, and physical function by MS subtype in a cross-sectional cohort (n=133). T scores are anchored to a general population mean of 50 with a standard deviation of 10. Higher scores indicate more of the trait being measured (e.g. depression scores above 50 indicate more depression than the average general population, physical function scores above 50 indicate more/better physical function than the average general population). RRMS: relapsing remitting MS; SPMS: secondary progressive MS; PPMS: primary progressive MS; Error bars: 95% confidence intervals.

Figure 3.

Mean PROMIS T scores for fatigue, pain, and physical function by self-reported disability in a cross-sectional cohort (n=133). T scores are anchored to a general population mean of 50 with a standard deviation of 10. Higher scores indicate more of the trait being measured (e.g. fatigue scores above 50 indicate more fatigue than the average general population, physical function scores above 50 indicate more/better physical function than the average general population). Error bars: 95% confidence intervals.

Table 1

Demographic data for all study participants (n=146), PROMIS completers (n=133), and non-completers (n=13).

	All Participants N (%)	PROMIS Completers N (%)	PROMIS Non-completers N (%)
N	146	133	13
Age (years, std)	51.16 ± 12.76	50.47 ± 12.80	58.15 ± 10.29
Female	113 (77)	104 (78)	4 (31)
Using Disease Modifying Therapy (DMT) *			
Yes	99 (68)	93 (70)	6 (46)
No	47 (32)	40 (30)	7 (54)
Education			
HS Diploma	55 (38)	77 (38)	4 (31)
College Grad	91 (62)	82 (62)	9 (69)
Type of MS			
Relapsing Remitting	104 (71)	99 (75)	5 (38)
Secondary Progressive	27 (19)	23 (17)	4 (31)
Primary Progressive	15 (10)	11 (8)	4 (31)
Disability **			
None/minimal	27 (19)	27 (20)	0 (0)
Mild	40 (27)	39 (30)	1 (8)
Moderate	35 (24)	32 (24)	3 (23)
Some support needed	29 (20)	24 (18)	5 (38)
Walker or two handed crutch	8 (5)	7 (5)	1 (8)
Unable to walk	7 (5)	4 (3)	3 (23)
Ethnicity			
Caucasian	133 (91)	122 (92)	11 (85)
Other	13 (9)	11 (8)	2 (15)

* DMTs used by participants included: glatiramer acetate, interferon-1b, interferon-1a, natalizumab, and fingolimod .

** Disability status was generated by a self-report scale previously shown to correlate with Expanded Disability Severity Scale scores ($r=0.85$) (20).

Table 2

Descriptive statistics for Legacy and PROMIS instruments.

Legacy Measures	N	Mean ± SD	Median	Min	Max	% Floor	% Ceiling	Alpha	No. items asked	Est. time to complete (sec)
Depression (BDI)	143	9.5 ± 8.4	7	0	44	10.5	0	0.91	21	252 [‡]
Anxiety (STAI)	146	40.5 ± 13.3	39	20	72	1.4	0	0.95	20	240 [‡]
Fatigue (MFIS)	120	40.7 ± 19.5	40	0	82	5	0	0.97	21	252 [‡]
Pain (PES)	139	13.6 ± 16.4	13	6	28	17.3	0	0.94	6	72 [‡]
Physical Function (SF-36)	142	38.04 ± 13.2	38.2	13.08	56.76	2.8	11.2	0.94 [*]	10	120 [‡]
PROMIS CAT Measures										
Depression 1.0	133	52.4 ± 9.2	51	34	78	-	-	0.98 [‡]	5	34 ± 17
Anxiety 1.0	133	55.1 ± 8.6	54	33	79	-	-	0.97 [‡]	4.5	35 ± 20
Fatigue 1.0	133	56.3 ± 9.5	58	28	76	-	-	0.99 [‡]	4.2	41 ± 18
Pain Interference 1.0	133	52.3 ± 9.4	53	39	72	-	-	0.85 [‡]	6.1	53 ± 59
Physical Function 1.0	133	42.5 ± 9.7	41	20	73	-	-	0.99 [‡]	4.5	84 ± 77

Abbreviations: BDI = Beck Depression Inventory; CAT = Computerized Adaptive Test; STAI = State Trait Anxiety Inventory; MFIS = Modified Fatigue Impact Scale; PES = Pain Effects Scale.

^{*} The Physical Function subscale of the SF-36 is scored on a T scale metric. Alpha for this scale was generated from raw data.

[‡] Alphas for PROMIS measures were obtained from the third party administrator AssessmentCenterSM.

[‡]Time to complete Legacy measures was estimated at 5 questions/minute (1).

Table 3

Comparison of Legacy and PROMIS scales. All scores converted to a T scale (referenced to the general population, mean of 50, standard deviation of 10).

Method/Trait		N	Mean	SD	SE	Median	Min	Max	p value*
Legacy	Depression	133	51.31	12.1	1.05	49.16	37.01	92.9	0.08
PROMIS	Depression	133	52.42	9.21	0.8	51	34	78	
Legacy	Anxiety	133	56.68	14.4	1.21	54.57	33.91	90.43	0.09
PROMIS	Anxiety	133	55.11	8.58	0.74	54	33	79	
Legacy	Physical Fn.	133	38.35	13.12	1.14	39.29	13.08	56.76	<.002
PROMIS	Physical Fn.	133	42.45	9.67	0.84	41	20	73	

*

Two-tailed pairwise t-test of the means.

Note: Legacy Fatigue and Pain scales used in this study are specifically designed for MS and therefore do not have normative data for the general US population that would allow conversion to a T metric.

Table 4
Multitrait-Multimethod Correlation Matrix comparing legacy and PROMIS questionnaires.

	PROMIS Measures					Legacy Measures				
	Dep	Anxiety	Fatigue	Pain	Phys Fn.	Dep	Anxiety	Fatigue	Pain	Phys Fn.
PROMIS										
Depression	<i>0.98</i>									
Anxiety	0.76	<i>0.97</i>								
Fatigue	0.58	0.52	<i>0.99</i>							
Pain	0.50	0.49	0.62	<i>0.85</i>						
Physical Function	-0.36	-0.19	-0.57	-0.53	<i>0.99</i>					
Legacy										
Depression	0.80	0.69	0.60	0.46	-0.35	<i>0.91</i>				
Anxiety	0.75	0.67	0.47	0.39	-0.23	0.80	<i>0.95</i>			
Fatigue	0.66	0.57	0.86	0.67	-0.62	0.68	0.61	<i>0.97</i>		
Pain	0.57	0.53	0.66	0.86	-0.54	0.60	0.47	0.74	<i>0.94</i>	
Physical Function	-0.29	-0.12	-0.42	-0.44	0.87	-0.27	-0.17	-0.52	-0.49	<i>0.94</i>