Arthritis Care & Research
Vol. 63, No. S11, November 2011, pp S86–S97
DOI 10.1002/acr.20531

Measures of Fibromyalgia

Fibromyalgia Impact Questionnaire (FIQ), Brief Pain Inventory (BPI), Multidimensional Fatigue Inventory (MFI-20), Medical Outcomes Study (MOS) Sleep Scale, and Multiple Ability Self-Report Questionnaire (MASQ)

DAVID A. WILLIAMS[1] AND LESLEY M. ARNOLD[2]

INTRODUCTION

The assessment of fibromyalgia (FM) is challenging because there are no biomarkers for this condition. Clinicians must rely upon patient-reported symptoms in order to understand the complexities of this condition. While in 1990, the American College of Rheumatology (ACR) developed research classification criteria involving tender point counts, it has only been within the past year that the ACR proposed clinical diagnostic criteria (1). Historically, many symptoms have been thought to be associated with FM. In order to narrow the field to those symptoms with the greatest clinical relevance, a working group within Outcome Measures in Rheumatology (OMERACT) conducted several Delphi exercises within both patients and clinicians to obtain consensus regarding which domains should be assessed in clinical trials for FM (2,3). The instruments to be reviewed herein reflect the clinically relevant domains defined by this OMERACT working group.

A wide variety of instruments have been used to index the OMERACT domains for FM. Many of the instruments were developed for use generically or have been borrowed from other clinical populations. In recent phase II and III clinical trials of medications for FM, wide variation was observed in the selection of domain indices (Table 1). While many of these measures are reviewed elsewhere in this special issue, we have selected a representative measure from each of the following domains of relevance: pain (Brief Pain Inventory), fatigue (Multidimensional Fatigue Inventory), sleep disturbance (Medical Outcomes Study Sleep Scale), and cognitive dysfunction (Multiple Ability Self-Report Questionnaire). Mood and functional status are also important domains for FM; however, the instruments most commonly used to assess these domains are reviewed elsewhere in this special issue and will not be repeated here (e.g., mood [Hospital Anxiety and Depression Scale] and functional status [Short Form 36]). Recent work in the development of responder indices suggests that either these specific instruments or other measurement tools from within the same domain can be used to differentiate responders from nonresponders in clinical treatment trials for FM (4). The precision by which these domains will be able to be assessed in the future is likely to be enhanced as newer measurements are being developed using either classic test construction methods or methods such as item response theory and computer adaptive testing, as is being done in the National Institutes of Health Patient-Reported Outcomes Measurement System (5).

Supported in part by the National Institute of Arthritis and Musculoskeletal and Skin Diseases/NIH (grants U01-AR55069 and R01-AR053207).

[1]David A. Williams, PhD: University of Michigan, Ann Arbor; [2]Lesley M. Arnold, MD: University of Cincinnati College of Medicine, Cincinnati, Ohio.

Dr. Williams has received consultancy fees, speaking fees, and/or honoraria (less than $10,000 each) from Bristol-Meyers Squibb, Forest Pharmaceuticals, Jazz Pharmaceuticals, and Lilly. Dr. Arnold has received consultancy fees (less than $10,000 each) from AstraZeneca, Cypress Bioscience, Eli Lilly, Forest Pharmaceuticals, Takeda, Sanofi-Aventis, Grünenthal Group, Wyeth, and Johnson & Johnson, and (more than $10,000) from Pfizer, and has received grants from Boehringer Ingelheim, Cypress Bioscience, Eli Lilly, Forest Pharmaceuticals, Novartis, and Pfizer.

Address correspondence to David A. Williams, PhD, Chronic Pain & Fatigue Research Center, University of Michigan, 24 Frank Lloyd Wright Drive, Lobby M, Ann Arbor, MI 48106. E-mail: daveawms@umich.edu.

Submitted for publication February 28, 2011; accepted in revised form May 10, 2011.

FIBROMYALGIA IMPACT QUESTIONNAIRE (FIQ)

Description

Purpose. The FIQ was developed in the late 1980s by clinicians at Oregon Health & Science University (OHSU) to assess the total spectrum of problems related to fibromyalgia (FM) and associated responses to therapy (6). The FIQ was first published in 1991 (7) and modified in both 1997 and 2002 to refine items and to clarify the scoring system (6). The FIQ was revised in 2009 (FIQR) to better reflect current understanding of FM and to address limitations of the original FIQ while retaining its essential properties (8).

Table 1. Outcome measures in fibromyalgia trials of Food and Drug Administration–approved medications

Fibromyalgia domain	Outcome measure
Pain	Visual analog scale (daily diary)
	Numeric rating scale (0–10) (daily diary)
	Fibromyalgia Impact Questionnaire pain (0–10)
	Brief Pain Inventory pain severity scores (0–10)
	Short Form 36 bodily pain
Tenderness	Dolorimetry (tender point threshold)
Fatigue	Visual analog scale (0–100) (daily diary)
	Fibromyalgia Impact Questionnaire fatigue (0–10)
	Short Form 36 vitality
	Multidimensional Fatigue Inventory
	Multidimensional Assessment of Fatigue
Sleep	Numeric rating scale (0–10) daily diary of sleep quality
	Fibromyalgia Impact Questionnaire morning rested feelings (0–10)
	Medical Outcomes Study sleep scale
Depression	Beck Depression Inventory
	Hamilton Depression Rating Scale
	Fibromyalgia Impact Questionnaire depression (0–10)
	Hospital Anxiety and Depression Scale depression
Anxiety	Fibromyalgia Impact Questionnaire anxiety
	Hospital Anxiety and Depression Scale anxiety
Cognition	Multiple Abilities Self-Report Questionnaire
Stiffness	Fibromyalgia Impact Questionnaire stiffness (0–10)
Physical function	Short Form 36 physical function
	Fibromyalgia Impact Questionnaire physical function

Content. The original FIQ (1991) covered 3 domains: function, overall impact, and symptoms. The function domain contained 10 physical functioning items related to the ability to perform large muscle tasks, including the ability to do shopping, do laundry, prepare meals, wash dishes by hand, vacuum a rug, make beds, walk several blocks, visit friends or relatives, do yard work, and drive a car. The overall impact domain contained 2 items asking about the number of days individuals felt well and the number of days they were unable to work because of FM symptoms. The symptoms domain contained 7 items using 10-cm visual analog scales on which patients rate work difficulties, pain, fatigue, morning tiredness, stiffness, anxiety, and depression. The 1997 version modified items about "work" to include "housework," and a new item about "climbing stairs" was added to the functioning domain. Finally, the 1997 version added hash marks (i.e., vertical lines) every 1 cm to the formatting of all visual analog scales. The 2009 FIQR has the same 3 domains as the original FIQ (function, overall impact, and symptoms), but differs in several ways. First, the physical functioning domain was reduced to 9 items and modified to reflect a better balance between large-muscle activities in the upper and lower extremities, and that would have less sex and ethnicity bias. The physical functioning items include the ability to brush or comb hair; walk continuously for 20 minutes; prepare a homemade meal; vacuum, scrub, or sweep floors; lift and carry a bag full of groceries; climb 1 flight of stairs; sit in a chair for 45 minutes; and go shopping for groceries. The overall impact domain was completely revised to reflect the overall impact of FM on functional ability and the overall impact of FM on the perception of reduced function. The symptom domain retained items on pain, fatigue, morning tiredness, stiffness, anxiety, and depression and added 4 additional items on tenderness, memory, balance, and environmental sensitivity.

Number of items. The original FIQ (1991) had 19 items capturing 3 domains. The 1997 version of the FIQ retained the same domains but added an additional item for a total of 20 items. In the 2009 FIQR, the first domain (physical function) has 9 items, the second domain (overall impact) has 2 items, and the third domain (symptoms) has 10 items for a total of 21 items.

Response options/scale. The physical functioning items in the 1991 and 1997 versions of the FIQ are rated on a 0–3 scale that best reflects the patient's ability to do the activity (0 = always, 1 = most, 2 = occasionally, 3 = never). The overall impact items are rated on a 0–7 scale for the number of days the patient felt well and the number of days the patient missed work, respectively. The symptom items are visual analog scales (0–10 cm), with higher numbers indicating greater symptomatology. All of the items in the 2009 FIQR are 0–10 numeric rating scales using 11 boxes, with higher numbers reflecting greater severity.

Recall period for items. The recall period is over the past week.

Endorsements/examples of use. Since 1991, the FIQ has been one of the most frequently used assessment tools in the evaluation of FM, and has been particularly useful as an outcome measure in FM clinical trials. The FIQ has been cited in over 300 articles between 1991 and 2010 (see URL: www.myalgia.com/FIQ/FIQ_REFS_2010.htm for a complete listing of article abstracts). The use of the FIQR in clinical studies has not yet been published.

Practical Application

How to obtain. The FIQ and the FIQR are free for academic and clinical use. An online license to use the FIQ is available by registering at URL: www.myalgia.com/FIQ/FIQ_academic_agreement.htm. The original FIQ is published in reference (7). The 1997 version with the 2002 scoring revision was published in 2005 (6) and is also available at URL: www.myalgia.com/FIQ/FIQ_B.htm. The FIQR is available at this same web site and was published in 2009 (8).

Method of administration. The FIQ and FIQR are administered as self-report questionnaires.

Scoring. The 1991 and 1997 FIQ versions have similar scoring. The final scores for each item of the FIQ should range from 0 (no impairment) to 10 (maximum impairment). The physical functioning items are rated on a 4-point Likert-type scale. Raw scores on each question can range from 0 (always) to 3 (never). Because some patients may not do some of the tasks listed, they are given the option of deleting questions from scoring. The scores for the items that the patient has rated are summed and divided by the number of questions answered. An average raw score between 0 and 3 is obtained. This value is then multiplied by 3.33. The first impact item that asks the number of days in the past week the patient felt well is reverse scored so that a higher number indicates impairment. Raw scores range from 0–7 and are then multiplied by 1.43. The second impact item is scored as the number of days the patient was unable to do regular work activities. Raw scores range from 0–7 and are then multiplied by 1.43. Symptom items are visual analog scales. In the 1991 version, the items are scored in number of cm from 0–10. Because the 1997 version added hash marks to all of the visual analog scales, these items are scored in numerical increments from 0–10, allowing scores to include 0.5 if the patient marks the space between 2 vertical lines. In the 1991 version, patients were instructed to cross out items 3 and 4 if they did not work. Therefore, the total maximum FIQ score was reduced from 100 to 80. With the 1997 revision in which questions 3 and 4 were modified to include housework, the total FIQ scores should always range from 0–100. In 2002, a modification of the scoring was recommended to address incomplete data. In order to maintain homogeneity on a 0–100 continuum, the final score is to be adjusted to reflect a final maximum score of 100. For example, if a patient missed 2 questions, the total recorded score should be adjusted by a factor of 10/8. The FIQR has 21 individual items and all items are based on an 11-point numeric rating scale of 0–10, with 10 being the "worst." The summed score for the function domain, which contains 9 items (range 0–90) is divided by 3; the summed score for overall impact, which contains 2 items (range 0–20) is not changed; and the summed score for symptoms, which contains 10 items (range 0–100) is divided by 2. As in the FIQ, the total maximum score for the FIQR is 100. The weighting of the 3 domains is different from the FIQ in that function accounts for 30% of the total score as opposed to 10% in the FIQ, the symptom domain makes up 50% of the score instead of 70% in the FIQ, and the overall impact domain remains the same as the FIQ at 20% (8).

Score interpretation. The final scores for each of the FIQ and FIQR items range from 0 (no impairment) to 10 (maximum impairment). The total maximum score for both the FIQ and the FIQR is 100, which represents the maximum impact of FM on the patient.

Respondent burden. It takes approximately 3–5 minutes to complete the FIQ. The FIQR is estimated to take just over 1 minute to complete.

Administrative burden. The FIQ and FIQR are easily administered by handing the questionnaires to the participant. The scales include simple instructions for the respondents. No formal training is required for the FIQ or FIQR. Scoring is relatively simple for both the FIQ and the FIQR but the use of numeric rating scoring for all of the FIQR items further simplifies the scoring and allows for use of electronic versions of the FIQR that can be administered online as was done in the validation study (8).

Translations/adaptations. The FIQ has been translated from English into 12 languages: Czech (Czech Republic), Dutch (The Netherlands), French (France and Canada), German (Germany), Hebrew (Israel), Italian (Italy), Korean (Korea), Polish (Poland), Romanian (Romania), Spanish (Argentina and Spain), Swedish (Sweden), Turkish (Turkey; see URL: www.myalgia.com/FIQ/FIQ_B.htm for more information on translations).

Psychometric Information

Method of development. The initial version of the FIQ was based on an intake questionnaire used by the OHSU rheumatology clinic and informal discussions with patients with FM. This FIQ was mailed at weekly intervals for a total of 6 weeks to 64 women with FM, along with the Arthritis Impact Measurement Scale (AIMS). A second group of 25 women with FM attending the OHSU Fibromyalgia Treatment Clinic completed the FIQ as part of their routine clinical evaluation. The construct validity, test–retest reliability, and content relevance of the FIQ were assessed in these 2 groups of patients (6,7). The FIQR was based on previous experience with the FIQ and patients' evaluation of important symptoms (8). The new questionnaire was tested in a focus group of 10 female patients with FM. Following discussions among the patients and investigators, agreement was reached on the final version of the FIQR. The FIQR was then tested in an online survey that was completed by patients with FM, rheumatoid arthritis (RA), systemic lupus erythematosus (SLE), or major depressive disorder (MDD), and healthy controls. The participants also completed the original FIQ and the 36-item Short Form Health Survey (SF-36).

Acceptability. The FIQ was originally developed to assess the current health status of women with FM, and may therefore have a sex bias, particularly in the functional items in which several of these questions relate to activities that are more likely to be performed by women. The functional questions were intended for a relatively affluent patient who was assumed to have possession of a car, a vacuum cleaner, and a washing machine and may therefore not generalize to all patients with FM. The FIQ also has problems related to the deletion of physical function items deemed "not applicable" by the respondent, which may result in an underestimation of functional severity. Some patients report difficulty understanding the scoring of the physical function questions and note that the questions do not allow them to rate the degree of difficulty in performing the activity. For example, a patient may report that they were "always" able to do shopping even though it took a great deal of time and effort to complete the task. The FIQ functional items are oriented toward high levels of disability, resulting in a potential floor effect. For example, in one study, 12% of patients scored a 0 on the FIQ physical function score (i.e., no dysfunction) (9). The FIQR

was developed to correct some of the problems with the FIQ. In particular, the physical functioning items were revised to have less sex and ethnicity bias than the FIQ and to improve the ease of scoring the functional activities on a 0–10 scale ranging from "no difficulty" to "very difficult" (8).

Reliability. In the original 1991 study to evaluate the FIQ, the test–retest reliability (Pearson's r) was assessed by the weekly recording of data over 6 weeks. The reliability ranged from 0.56 on the pain score to 0.95 for physical function (7). The internal consistency (Cronbach's alpha) was not reported in the original analysis. The Cronbach's alpha for the FIQR was 0.95, with item-total correlations ranging from 0.56–0.93. Test–retest reliability was not determined for the FIQR (8).

Validity. The content validity of the original FIQ was assessed from an analysis of missing data for each item. Missing data from the physical functioning items were limited to 11% of patients who did not do dishes by hand and 20% who did no yard work. Because many patients were not working outside the home, the 2 work items of the original FIQ were not relevant for 38% of the patients (6,7). In the validation study of the FIQR, patient suggestions about content and wording of the instrument during the focus group meeting contributed to the face validity of the final version of the FIQR. Content validity of the FIQR was suggested by strong correlation between the FIQR and the SF-36. For example, the FIQR function domain was most highly correlated with the SF-36 physical functioning subscale (8). The construct validity of the 1991 FIQ was determined by measuring the correlation of the FIQ individual items with the AIMS. The FIQ physical functioning items had a significant correlation (r = 0.67) with the AIMS lower-extremity physical function component score. The pain, depression, and anxiety items of the FIQ showed significant correlations with the corresponding AIMS scales (0.69, 0.73, and 0.76, respectively). The AIMS visual analog of syndrome impact correlated least robustly with the FIQ items, the highest correlation being with pain (r = 0.48). Item correlations with the AIMS syndrome activity question tended to be higher, ranging from 0.28–0.83. A principal components analysis yielded 5 factors. The 10 physical functioning questions loaded on the first factor with component loading ranging from 0.50 to 0.95. Factor 2 consisted of work difficulty, feeling good, pain, fatigue, rest, and stiffness. Anxiety, depression, and days of work missed all loaded on separate factors (6,7). Convergent validity was assessed by comparing the FIQR to both the SF-36 and the FIQ. The 3 domains of the FIQR and the associated individual items correlated closely with the corresponding subscales on the SF-36. Each of the 3 FIQR domains was also highly correlated with the total FIQR score. There was a strong correlation (0.88) between the FIQR and the FIQ, suggesting that the questionnaires are capturing similar information about the impact of FM. The mean total score of the FIQR was ~4 points lower than the mean FIQ total score, which was attributed to the change of the weighting in the FIQR scoring (8). Each of the 3 FIQR domains predicted unique variance in SF-36 domains, providing evidence for discriminant validity. Discriminant validity was also evaluated by comparing the FIQR total scores in patients with FM with the scores in healthy controls, patients with RA or SLE, and patients with MDD. The FM FIQR scores were significantly higher than in the other 3 groups (8).

Ability to detect change. The FIQ has been most commonly used as an outcome measure in treatment trials and, in general, has demonstrated an ability to detect clinical change (6). The FIQ total score was also included as an outcome measure in trials of the 3 US Food and Drug Administration–approved medications for FM, pregabalin, duloxetine, and milnacipran (10–12). For example, in a pooled analysis of 4 placebo-controlled, double-blind studies of duloxetine in FM, the total FIQ scores improved significantly in the duloxetine groups compared with placebo, with a mean (SE) reduction of 12.62 (0.61) in the duloxetine patients compared with a mean (SE) reduction of 8.2 (0.69) in the placebo group ($P < 0.001$) (13). A recent study suggested that a 14% change or an absolute change of 8.1 (95% confidence interval 7.6–8.5) in the FIQ total score represented a clinically meaningful change in FM status (i.e., minimum clinically important difference). The minimum clinically important difference was determined by calculating the percentage change in the FIQ total score from baseline and linking this to each patient's global assessment of change score (14).

References. The validation of the original FIQ is published in an article by Burckhardt et al (7). A review of the development, operating characteristics, and uses of the FIQ was done by Bennett (6) and the validation study of the FIQR is found in the Bennett et al publication in 2009 (8).

Critical Appraisal of Overall Value to the Rheumatology Community

Strengths. FM is associated with multiple symptoms and functional impairment. The FIQ and FIQR are useful assessment tools in FM because they evaluate the total spectrum of problems related to FM, including functional impairment, overall impact, and FM-related symptoms. The FIQ total score has proved to be a useful outcome measure in key clinical trials of FM.

Caveats and cautions. The FIQ functional items are oriented toward high levels of disability, resulting in a possible floor effect. Because the FIQ was originally developed in a patient population of relatively affluent women, there is a potential problem with sex and ethnicity bias. Although the individual domains and/or items on the FIQ were not originally intended to be used in isolation, some recent studies have reported single-item or domain scores from this instrument. The internal consistency (Cronbach's alpha) was not reported in the original analysis of the FIQ. The FIQR was designed to correct some of the problems with the FIQ, but has not yet been tested in the context of clinical trials. Test–retest reliability was not determined for the FIQR.

Clinical usability. The FIQ and FIQR are brief, self-report questionnaires that assess the impact of FM on patients. The FIQ has most commonly been used in clinical studies, but has the potential for use in the clinical setting to monitor patients' response to treatment over time.

Research usability. The FIQ has been used in large-scale clinical trials of therapeutics for FM, supporting its ability to assess and detect change in FM.

BRIEF PAIN INVENTORY (BPI)

Description

Purpose. The BPI was designed to measure multiple clinically relevant aspects of pain such as pain intensity and interference from pain in cancer populations (15). The BPI was originally called the Wisconsin Brief Pain Questionnaire (16). Subsequently, support for its valid use in noncancer populations such as musculoskeletal, neuropathic, and other central pain conditions has been established (17,18). There are 2 versions; the short version is the most commonly used and is often included in the context of clinical trials. This is the version that possesses most foreign language translations. A longer, less frequently used version is available that includes more pain descriptors and may have clinical utility; however, the developers recommend the short form for most applications. Only the shorter form will be considered here.

Content. The BPI assesses for the presence of pain, pain intensity (i.e., worse, least, average, current), and functional interference from pain (i.e., activity, mood, walking ability, normal work, relations with others, sleep, life enjoyment). It also catalogs the types of pain medications being used, the percentage of pain relief obtained from medications, and assesses the distribution of pain via a body map.

Number of items. The BPI contains a total of 15 items.

Response options/scale. The BPI uses a mixture of item types. Item 1 querying about the presence of pain is a dichotomous "yes," "no." Item 2, the body map, asks that areas of pain be shaded and an "X" placed on the body region that hurts the most. Items 3–6 (intensity items) utilize a 0 (no pain) to 10 (pain as bad as you can imagine) 11-point rating scale. Item 7 is an open-ended response to list pain medications. Item 8 (percentage of pain relief) ranges from 0% (no relief) to 100% (complete relief). Item 9 (a–g) inquires about interference using an 11-point numeric rating scale. Each item ranges between 0 (does not interfere) and 10 (completely interferes).

Recall period for items. The time frame for the BPI is typically based upon "the past week" but some versions allow for the past 24 hours.

Endorsements/examples of use. The BPI is widely used in clinical trials for pain and in pain research generally. It is one of the instruments recommended by the Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials group (19) for inclusion in any clinical trial evaluating pain.

Practical Application

How to obtain. The BPI is available through the following address: The Department of Symptom Research, Attn: Assessment Tools, The University of Texas MD Anderson Cancer Center, 1515 Holcombe Boulevard, Unit 1450, Houston, TX 77030. Phone: 713-745-3805. The BPI is available free of charge for nonfunded academic research. For funded academic research there is a charge per project (e.g., $300) and a charge for commercial research (e.g., $800 per project).

Method of administration. The BPI can be administered as a self-report questionnaire or as an interview.

Scoring. While some of the items represent single-item values, pain intensity, indexed by the Pain Severity Score, is calculated by obtaining the mean of the 4 pain intensity items. The Pain Interference Score is obtained by calculating the mean of the 7 interference items.

Score interpretation. The Pain Severity Score has a maximum value of 10 (i.e., "pain as bad as you can imagine") and a minimum value of 0 (i.e., "no pain"). The Pain Interference Scale similarly has a maximum value of 10 (i.e., "completely interferes") to 0 (i.e., "does not interfere"). The BPI is easily scored by hand.

Respondent burden. It takes approximately 5 minutes to complete the BPI.

Administrative burden. The BPI is easily administered by handing the questionnaire to the participant or by asking each question verbally. Scoring is accomplished by calculating 2 means, which can be done in <5 minutes.

Translations/adaptations. Validated translations are available for the following languages: English, Spanish, Italian, Russian, Norwegian, Greek, German, Japanese, Chinese, Arabic, Bulgarian, Cebuano, Croatian, Czech, Filipino, French, Hindi, Korean, Malay, Slovak, Slovenian, and Thai.

Psychometric Information

Method of development. Prior to the development of the BPI, there was no specific instrument designed to the intensity and impact of cancer pain that was brief and that could be administered repeatedly over time to monitor the effects of treatment. Existing measures at the time (e.g., the McGill Pain Questionnaire) were developed for noncancer pain. Based upon patient interviews, it was discovered that existing questionnaires were too ambiguous, irrelevant, or too lengthy for the assessment of cancer pain. The questionnaire was developed in accordance with the best guidelines for test construction available at the time (i.e., the 1970s; *Standards for Educational and Psychological Tests* published by the American Psychological Association, American Educational Research Association, and by the National Council on Measurement in Education). Item development was informed by patient interviews and by field testing of items. Even though this questionnaire was developed 30 years ago, the approach conforms to the more recently published *Draft Guidance for Industry, Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims by the FDA.* The BPI has since been validated for use as a brief

and meaningful pain assessment tool for noncancer pain conditions as well (17,18).

Acceptability. Acceptability was assessed in a noncancer pain population. The BPI was readily accepted by patients, was not associated with excessive missing data, and did not have problematic floor/ceiling effects (20).

Reliability. Internal consistency for the Pain Severity Score and for the Interference scale has been reported as being 0.85 and 0.88, respectively, in noncancer pain populations (18). Test–retest reliability has been assessed for both cancer and noncancer forms of pain and for over varying time frames. For very short time intervals (e.g., 30–60 minutes), the test–retest reliability was 0.98 for pain severity and 0.97 for pain interference (21). Test–retest reliability for daily administration ranges between 0.83–0.88 for pain severity and between 0.83–0.93 for pain interference (22). FM is considered to be a form of noncancer or musculoskeletal pain and as such these metrics could be applied to FM; however, formal assessment of reliability of the BPI in FM is not available.

Validity. Item analysis has consistently revealed a 2-factor structure (severity or intensity and interference) in more than 36 studies of the BPI across multiple languages for both cancer and noncancer pain populations (23). Construct validity of the BPI has been supported for the generic assessment of pain as well as specifically for low back pain, rheumatoid arthritis (17), and osteoarthritis (20). In a sample of patients with arthritis, the BPI pain severity score correlated ($r = 0.74$) with the bodily pain scale of the Short Form 36, a generic measure of pain intensity, and ($r = 0.77$) with the Chronic Pain Grade Intensity scale, another generic pain intensity measure. The BPI Interference scale from this same sample correlated ($r = 0.81$) with the Chronic Pain Grade disability scale, and ($r = 0.69$) with the Health Assessment Questionnaire disability index, a disease-specific measure of functional interference (17).

Ability to detect change. The BPI has demonstrated response to change in response to many forms of pharmacologic and nonpharmacologic treatments (23). In chronic pain states, generally, an improvement of 30% or 2–3 points improvement is considered to be a clinically meaningful change (24–26). In a pooled analysis across 12 weeks of treatment from 4 randomized controlled trials of duloxetine for fibromyalgia (FM), the BPI "average pain" and the "Pain Severity Score" was anchored against the Patient Global Impression of Improvement scale (PGI-I). Anchor-based minimum clinically important differences for the "average" pain and for the PGI-I were calculated based upon the difference in mean change from baseline to end point resulting in values of 2.1 and 2.2 points, respectively. This amount of change was associated with 32% and 34% reductions in pain from the baseline scores, respectively (27).

References. The user manual for the BPI contains a reference listing of 72 studies supporting the valid use of the BPI across a wide variety of chronic pain conditions, including FM (23).

Critical Appraisal of Overall Value to the Rheumatology Community

Strengths. The BPI was designed to monitor change in pain (and its impact) over time. Numerous studies support its validity to function in this capacity.

Caveats and cautions. The BPI is an industry standard for the generic assessment of both cancer and noncancer pain conditions and contains few flaws in terms of psychometrics, ease of administration, or utility. Far more is known about the psychometrics of the Pain Severity scale and the Pain Interference scale than about the other features of the questionnaire (pain relief, body map, etc.). These other features are often not reported in trials using this instrument. Reports specifically focused upon the psychometric evaluation of the BPI in FM are not available; however, FM is classified as a chronic noncancer musculoskeletal pain condition and the validity of the BPI is supported for the generic assessment of pain intensity and interference.

Clinical usability. The BPI is recommended for use in clinical settings to monitor the severity and impact of pain generically.

Research usability. The BPI is recommended as tool of choice for the assessment of pain in clinical pain trials (28). It is easily administered and has low patient burden.

MULTIDIMENSIONAL FATIGUE INVENTORY (MFI-20)

Description

Purpose. The MFI-20 was introduced 1995 (29) as a measure of fatigue severity. Fatigue is perhaps the most common complaint heard by clinicians. Apart from the everyday use of the term to describe normal tiredness, it can be used to indicate the presence of disease (29). Therefore, the MFI-20 was developed to function as an index of disease, as a diagnostic criterion, or as an outcome variable when a treatment is being evaluated.

Content. The MFI-20 possesses 5-factor analytically confirmed subscales assessing general fatigue, physical fatigue, reduced activity, reduced motivation, and mental fatigue. The MFI differs from other multidimensional fatigue measures by purposely retaining a relatively short list of items, and by eliminating somatic items.

Number of items. The MFI-20 contains 20 items.

Response options/scale. The MFI-20 uses the same response set for each of the 20 items. The respondent is asked to mark an X in 1 of 5 boxes arranged linearly and anchored by "yes, that is true" at one pole to "no, that is not true" at the opposite pole. Scoring of scales requires some items to be reversed such that a higher score on each scale is indicative of greater fatigue.

Recall period for items. The time frame is somewhat nonspecific as the questionnaire queries for symptoms occurring "lately."

Endorsements/examples of use. The MFI-20 has been used in numerous clinical populations, including cancer (30), Sjögren's syndrome (31), craniopharyngioma (32), myelodysplastic patients (33), chronic fatigue syndrome

(29), fibromyalgia (FM) (34), and general chronic pain (35). It has also been validated for use in nonclinical samples, including psychology students, medical students, Army recruits, and junior physicians (29).

Practical Application

How to obtain. The MFI-20 is available from the author: E. M. A. Smets, Academic Medical Centre, University of Amsterdam, Department of Medical Psychology, Amsterdam, The Netherlands.

Method of administration. The MFI-20 is a self-report questionnaire.

Scoring. Each scale can be calculated by summing the specific items within each scale. Some items need to be reverse scored prior to summing.

Score interpretation. Each scale contains 4 items with a maximum value of 20 (i.e., each item is endorsed with a "5") and a minimum value of 4 (i.e., each item is endorsed with a "1"). Higher scores on each scale indicate more fatigue severity.

Respondent burden. It takes approximately 5 minutes to complete the MFI-20.

Administrative burden. The MFI-20 is easily administered by handing the questionnaire to the participant. Scoring is accomplished by reverse scoring the required items and then summing each of the 5 scales. Scoring can be completed in <5 minutes.

Translations/adaptations. Validated translations are available for the following languages: English, Swedish, French, and German.

Psychometric Information

Method of development. At the time of development, both 1-dimensional and multidimensional measures of fatigue existed but were quite lengthy and confounded by somatic items. With a consideration of the legacy measures of the time, development of the MFI was initiated by postulating the existence of 5 dimensions of fatigue. Items were generated and then field tested in a diverse group of individuals expected to experience a wide range of fatigue, including individuals with cancer, individuals with chronic fatigue syndrome, first-year medical and psychology students, junior physicians, and Army recruits. Confirmatory factor analysis supported the retention of the 5-dimensional model inherent in this instrument (29).

Acceptability. The MFI is not associated with excessive missing data problems or with or floor/ceiling effects (36).

Reliability. In the original validation study, internal consistency (Cronbach's alpha) for the 5 scales ranged between 0.53–0.93 with the average being 0.80 (29). A more recent validation study of the MFI-20 conducted in the US with a general population sample found the following Cronbach's alpha values: general fatigue (0.83), physical fatigue (0.81), reduced activity (0.82), reduced motivation (0.71), and mental fatigue (0.86) (36). Internal consistency of a total of all 20 items was 0.93. Test–retest reliability has not been reported.

Validity. Confirmatory factor analysis has repeatedly found a 5-factor solution as best fitting the data (i.e., general fatigue, physical fatigue, reduced motivation, reduced activity, mental fatigue), each with adjusted goodness of fit indexes above 0.90 (30). Convergent validity was supported by comparing each scale to a visual analog scale (VAS) assessing fatigue. Associations were all significant with the general fatigue scale having the strongest relationship (30). Construct validity for each scale in association with other relevant constructs has been supported in several validation studies for the MFI-20 (29,30,36).

Ability to detect change. Formally established minimum clinically important differences have not been published for the MFI-20 in FM, however each of the scales appear to be responsive to treatment changes, especially the general fatigue scale (30).

References. There is no specific user manual but the original manuscript provides details on the development and psychometrics of the instrument (29).

Critical Appraisal of Overall Value to the Rheumatology Community

Strengths. The MFI-20 is a brief measure of fatigue that appears to capture relevant dimensions of fatigue severity. It has been used successfully in FM and appears to be a good marker of illness across a broad range of medical illnesses. While not as brief as a single-item VAS (as is commonly used), the MFI-20 correlates well with these measures but offers greater clarification of the type of fatigue being experienced and offers better assessment precision than single-item measures. The MFI does a good job of capturing the experience of fatigue across multiple dimensions without being contaminated by constructs such as functional status (i.e., the functional impact of fatigue), which is better assessed by functional status measures.

Caveats and cautions. Five levels of "yes, that is true" to "no, that is not true" represent a difficult response set for some patients to interpret.

Clinical usability. The MFI-20 may be too lengthy for the typical clinic where a briefer screen may be more appropriate. If however there is a desire to track specific forms of fatigue over time, then this is an appropriate measure.

Research usability. The MFI-20 is recommended for use in clinical trials of interventions targeting fatigue. It has been used successfully in clinical trials of FM (37).

MEDICAL OUTCOMES STUDY (MOS) SLEEP SCALE

Description

Purpose. The MOS Sleep Scale was originally developed as part of the MOS, which was a 4-year observational study of health outcomes for chronically ill patients. The MOS Sleep Scale represents the portion of this larger assessment protocol that specifically focused upon sleep (38). The MOS Sleep Scale is a non–disease-specific measure of multiple aspects of sleep problems.

Content. The MOS Sleep Scale is a 12-item measure assessing 6 domains of sleep: 1) sleep disturbance (e.g., the

ability to fall and stay asleep), 2) sleep adequacy (e.g., sleeping enough to feel rested and restored), 3) sleep quantity (e.g., the number of hours slept), 4) somnolence (e.g., daytime sleepiness), 5) snoring, and 6) shortness of breath or headache.

Number of items. The MOS Sleep Scale contains 12 items in its original form; this form has been used in the context of fibromyalgia (FM) clinical trials (37,39) and will be the focus of this review. A briefer 6-item version is also available from the publisher.

Response options/scale. The MOS Sleep Scale uses a variety of response sets. Item 1 queries about how long it takes to fall asleep. Response options are blocked into "0–15 minutes," "16–30 minutes," "31–45 minutes," "46–60 minutes," and "more than 60 minutes." Item 2 queries about how many hours of sleep were obtained on average over the past 4 weeks. This is an open-ended question ranging between 0–24 hours. The remaining 10 items use a 6-point response set based upon the following values and anchors (1 = all of the time, 2 = most of the time, 3 = a good bit of the time, 4 = some of the time, 5 = a little of the time, and 6 = none of the time).

Recall period for items. The time frame for each item is the past 4 weeks. An acute 1-week recall version is also available.

Endorsements/examples of use. The MOS Sleep Scale has been used in numerous nonclinical and clinical populations, including a general US sample (40), cancer pain (41), restless legs syndrome (42), overactive bladder (43), rheumatoid arthritis (44), dialysis (45), neuropathic pain (46), and FM (47).

Practical Application

How to obtain. The MOS Sleep Scale is available from its publisher, Quality Metric. More information can be found at URL: QualityMetric.com. It is recommended that the interested user contact the publisher to learn about potential pricing or licensing agreements associated with the use of this instrument.

Method of administration. The MOS Sleep Scale is a self-report questionnaire.

Scoring and score interpretation. Each scale can be hand scored. Some scales are single items and do not require scoring while others require items to be reversed and summed. Each scale (except sleep quantity) is recalibrated onto a 0–100 scale. For most scales, higher scores indicate worse sleep problems. The exceptions are sleep adequacy and sleep quantity where lower scores indicate worse sleep problems. The MOS Sleep Scale can be aggregated to produce 2 summary indices, the Sleep Problems Index II (9 items) and the Sleep Problems Index I (6 items). Each of these indices integrates the domains of sleep disturbance, sleep adequacy, shortness of breath, and somnolence into a single score. The difference between Sleep Problems Index 1 and 2 is simply length rather than domain coverage; potentially overlapping items were eliminated in Index 1. Higher scores on either index are indicative of worse sleep problems.

Respondent burden. It takes approximately 3–5 minutes to complete the MOS Sleep Scale.

Administrative burden. The MOS Sleep Scale is easily administered by handing the questionnaire to the participant. Scoring requires some reverse scoring, recalibrating scales onto a 0–100 scale, and aggregating the 2 summary indices. It can take 5–7 minutes to score.

Translations/adaptations. The 12-item version is available in 85 languages, which are available from the publisher.

Psychometric Information

Method of development. The MOS Sleep Scale was developed using an extensive review of the published sleep literature resulting in the selection of the domains contained in the scaling of this instrument. The intent was to construct an instrument that would identify sleep problems across sleep-related diseases and associated illnesses rather than being specific to any one type of problem. The scale was initially field tested in a large sample of healthy individuals as well as individuals with a variety of chronic illnesses associated with the MOS (42).

Acceptability. In an evaluation of the MOS Sleep Scale in neuropathic pain, missing items were observed in <10% of the sample. Ceiling and floor effects for each item were acceptable (i.e., <0.50% of all cases). A single item, "awakening short of breath," accounted for much of the problems in scaling properties (46). A second study found similar characteristics for a restless legs syndrome sample with <5% of cases experiencing floor or ceiling effects for the scale as a whole and <20% experiencing floor or ceiling effects for summed scales and <50% for individual items (42).

Reliability. Taken from the neuropathic pain study above (46), Cronbach's alpha ranged between 0.64–0.84 for the MOS sleep subscales. In restless legs syndrome all scales exceeded Cronbach's alpha of 0.70 with the exception of somnolence ($\alpha = 0.66$) (42). In a study of FM all multi-item scales (i.e., sleep disturbance, sleep adequacy, somnolence, and summary indices) exceeded $\alpha = 0.70$ (47).

Validity. Support for construct validity was identified in the restless legs syndrome study where worsening MOS sleep domain scores correlated strongly with worsening indices of quality of life (42). Multitrait scaling was used in the neuropathic pain sample to support convergent and divergent construct validity (46) and recently, confirmatory factor analysis has supported the factorial structure of the MOS Sleep Scale in FM (47). Qualitative interviews (i.e., cognitive debriefing) with patients with FM demonstrated that the MOS Sleep Scale was of relevance to individuals with FM and adequately captured the experience of sleep difficulties arising in FM (48). Additional work associated with criterion validity is needed for the MOS Sleep Scale when specifically applied to FM.

Ability to detect change. In a neuropathic pain sample, the minimal important difference for the 9-item Problem Index 2 was 5.1 (scale 0–100) (46). This is considered a moderate effect (0.65) and corresponds to the corrected change in a group of patients demonstrating change contrasted to the variation observed in a group of patients demonstrating no change. A study in FM reported a clin-

ically important difference (CID) for the sleep disturbance subscale as being 7.9 points (47). CID was calculated by examining differences from baseline as a function (i.e., anchored) of the Patient Global Impression of Change.

References. The publisher, Quality Metric, provides references regarding the development and psychometrics of this instrument.

Critical Appraisal of Overall Value to the Rheumatology Community

Strengths. The MOS Sleep Scale is widely used and is a generic measure of sleep problems that can be used to compare different clinical populations to one another on a common metric. The questionnaire is brief, responsive to change, and has been used in FM.

Caveats and cautions. The items do not use a uniform structure and the scoring is relatively complex given its brevity. The interpretation of the 2 composite indices is not completely obvious except that they are a combination of the assessed domains. Additional data supporting validity and responsiveness to change in FM are desirable.

Clinical usability. The MOS Sleep Scale can be used clinically to monitor changes in sleep across time and within broadly based domains of sleep problems; however, it is a bit lengthy for routine clinical use (48).

Research usability. The MOS Sleep Scale can be used to monitor treatment effects and appears to be sensitive to change both in sleep and in overall quality of life when sleep or other related symptoms improve or worsen.

MULTIPLE ABILITY SELF-REPORT QUESTIONNAIRE (MASQ)

Description

Purpose. The MASQ was purposely designed to assess the self-perception of cognitive difficulties in contrast to the more traditional "objective" neuropsychological assessment by a clinician (49). At the time of development, there were several measures of perceived memory problems, but other relevant areas of cognition lacked a valid self-appraisal tool.

Content. The MASQ contains items about perceived cognitive difficulties in 5 domains of clinical neuropsychological evaluation. The domains of the MASQ along with neuropsychological tests commonly used to index each domain are (50) language (L): Boston Naming Test, Controlled Oral Word Association (C, F, and L words and animals); visual-perceptual ability (VP): Wechsler Adult Intelligence Test Revised (WAIS-R; Block Design, Judgment of Line Orientation); verbal memory (VM): California Verbal Learning Test (Trials 1–5 total, Long Delay Free Recall), Wechsler Memory Scale Revised (WMS-R; Logical Memory I and II); visual-spatial memory (VSM): Rey-Osterrieth Complex Figure (immediate and delayed reproduction), WMS-R Visual Reproduction I and II; and attention/concentration (AC): Stroop Color-Word Test, WAIS-R Arithmetic, WAIS-R Digit Span.

Number of items. The MASQ contains 38 items.

Response options/scale. The MASQ uses the same 5-point response set for all items verbally anchored by "never," "rarely," "sometimes," "usually," and "always." The 5 scales (i.e., L, VP, VM, VSM, AC) are summed. A total score is produced by combining all items.

Recall period for items. No time frame is indicated on the original form.

Endorsements/examples of use. The MASQ has been used to assess perceived cognitive problems in several populations, including the following: epilepsy (49–51), adjuvant chemotherapy for breast cancer (52), breast cancer survivors (53), and fibromyalgia (FM).

Practical Application

How to obtain. The MASQ is available through the instrument's author: Michael Seidenberg, Department of Psychology, UHS/CMS, 3333 Green Bay Road, North Chicago, IL 60064.

Method of administration. The MASQ is administered as a self-report questionnaire.

Scoring. Each item is scaled between 1–5. Nearly half of the items require reverse scoring prior summing. Each scale is then summed. A total score containing all items is also possible. The maximum score for the total score is 190 (i.e., 38 items $\times$ 5). Scales containing 8 items (i.e., L, VM, VSM, AC) have a maximum score of 40 and VP (6 items) has a maximum score of 30.

Score interpretation. Higher scores on any scale indicate greater perceived difficulties with that cognitive domain.

Respondent burden. It takes approximately 10 minutes to complete the MASQ.

Administrative burden. The MASQ is easily administered by handing the questionnaire to the participant. Scoring is relatively simple but does require reverse scoring for nearly half of the items before summing.

Translations/adaptations. The MASQ is available in English.

Psychometric Information

Method of development. The initial version of the MASQ contained 48 items based upon clinical experience and a review of published questionnaires at the time of development. Content relevance was evaluated by 8 clinical neuropsychologists and 1 neuropsychiatrist with respect to the cognitive function depicted by each item. Agreement among raters for the retained items supports the content validity of each item.

Acceptability. In the development sample, 22% missed at least 1 item. Ceiling and floor effects were not reported.

Reliability. In the original validation sample, Cronbach's alpha was 0.92 for the total score. Internal consistency was above 0.70 for each of the individual scales (49). In other clinical samples, similar reliability estimates have been reported (e.g., $\alpha = 0.93$ for total and ranged from 0.72–0.79 for subscales in breast cancer survivors) (53). In the original validation study, 2-month test–retest reliability for the entire questionnaire was 0.71 and ranged

between 0.55 (L) and 0.74 (VM) (49). Test–retest data and internal consistency data is not available for FM.

Validity. In the original development of the MASQ, items were field tested in 2 samples, individuals with unilateral temporal-lobe epilepsy and healthy normal individuals. Support for concurrent validity came from higher MASQ scores being associated with poorer performance on neuropsychological tests in both samples but with greater perceived difficulties being observed in the clinical sample. These studies support the idea that perceived cognitive difficulties correspond to more objectively assessed indices of the same constructs (49). In a study comparing individuals with FM to healthy controls, individuals with FM scored significantly higher on each MASQ subscale than did healthy controls (54). Studies assessing the criterion validity of the MASQ with objective neuropsychological performance tests in FM are not currently available.

Ability to detect change. Reliable change indices and standard regression-based change norms have been established for the MASQ for use in cases of epilepsy (51). The MASQ has also demonstrated response to change in clinical trials of therapeutics for FM (e.g., milnaciparan) (55).

References. Original support for the MASQ is found in the work by Seidenberg et al (49).

Critical Appraisal of Overall Value to the Rheumatology Community

Strengths. Fibro fog is a common complaint among individuals with FM. Often only the memory aspects are assessed, but patients complain of broader deficits that are covered by the MASQ. The MASQ can be useful in tracking the varied manifestations of dyscognition in FM that are related to the different symptoms that characterize FM.

Caveats and cautions. The length of this instrument at 38 items may be prohibitive in settings where multiple domains of clinical relevance need to be efficiently measured. The MASQ has not been as rigorously developed or tested as the other measures reviewed in this article, but is one of the few measures currently available to assess this important aspect of FM.

Clinical usability. The MASQ appears to capture multiple aspects of fibro fog. Patients express a desire to have this domain assessed; yet, there are few instruments aside from the MASQ that are available for this purpose.

Research usability. The MASQ has been used in several large scale clinical trials of therapeutics for FM supporting is ability to assess and detect change in perceived cognitive difficulties.

AUTHOR CONTRIBUTIONS

Both authors were involved in drafting the article or revising it critically for important intellectual content, and both authors approved the final version to be published.

REFERENCES

1. Wolfe F, Clauw DJ, Fitzcharles MA, Goldenberg DL, Katz RS, Mease P, et al. The American College of Rheumatology preliminary diagnostic criteria for fibromyalgia and measurement of symptom severity. Arthritis Care Res (Hoboken) 2010;62:600–10.
2. Arnold LM, Crofford LJ, Mease PJ, Burgess SM, Palmer SC, Abetz L, et al. Patient perspectives on the impact of fibromyalgia. Patient Educ Couns 2008;73:114–20.
3. Mease PJ, Arnold LM, Crofford LJ, Williams DA, Russell IJ, Humphrey L, et al. Identifying the clinical domains of fibromyalgia: contributions from clinician and patient Delphi exercises. Arthritis Rheum 2008;59:952–60.
4. Arnold LM, Mease PJ, Williams DA, Martin SA, Wang F, Emir B, et al. Development of responder definitions for fibromyalgia clinical trials. Arthritis Rheum 2010;62:S38.
5. 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 2010;63:1179–94.
6. Bennett RM. The Fibromyalgia Impact Questionnaire (FIQ): a review of its development, current version, operating characteristics and uses. Clin Exp Rheumatol 2005;23:S154–62.
7. Burckhardt CS, Clark SR, Bennett RM. The Fibromyalgia Impact Questionnaire: development and validation. J Rheumatol 1991;18:728–33.
8. Bennett RM, Friend R, Jones KD, Ward R, Han BK, Ross RL. The Revised Fibromyalgia Impact Questionnaire (FIQR): validation and psychometric properties. Arthritis Res Ther 2009;11:R120.
9. Wolfe F, Hawley DJ, Goldenberg DL, Russell IJ, Buskila D, Neumann L. The assessment of functional impairment in fibromyalgia (FM): Rasch analyses of 5 functional scales and the development of the FM Health Assessment Questionnaire. J Rheumatol 2000;27:1989–99.
10. Arnold LM, Rosen A, Pritchett YL, D'Souza DN, Goldstein DJ, Iyengar S, et al. A randomized, double-blind, placebo-controlled trial of duloxetine in the treatment of women with fibromyalgia with or without major depressive disorder. Pain 2005;119:5–15.
11. Arnold LM, Russell IJ, Diri EW, Duan WR, Young JP Jr, Sharma U. A 14-week, randomized, double-blinded, placebo-controlled monotherapy trial of pregabalin in patients with fibromyalgia. J Pain 2008;9:792–805.
12. Arnold LM, Gendreau RM, Palmer RH, Gendreau JF, Wang Y. Efficacy and safety of milnacipran 100 mg/day in patients with fibromyalgia: results of a randomized, double-blind, placebo-controlled trial. Arthritis Rheum 2010;62:2745–56.
13. Arnold LM, Clauw DJ, Wohlreich MM, Wang F, Ahl J, Gaynor PJ, et al. Efficacy of duloxetine in patients with fibromyalgia: pooled analysis of 4 placebo-controlled clinical trials. Prim Care Companion J Clin Psychiatry 2009;11:237–44.
14. Bennett RM, Bushmakin AG, Cappelleri JC, Zlateva G, Sadosky AB. Minimal clinically important difference in the fibromyalgia impact questionnaire. J Rheumatol 2009;36:1304–11.
15. Cleeland CS. Measurement and prevalence of pain in cancer. Semin Oncol Nurs 1985;1:87–92.
16. Daut RL, Cleeland CS, Flanery RC. Development of the Wisconsin Brief Pain Questionnaire to assess pain in cancer and other diseases. Pain 1983;17:197–210.
17. Keller S, Bann CM, Dodd SL, Schein J, Mendoza TR, Cleeland CS. Validity of the brief pain inventory for use in documenting the outcomes of patients with noncancer pain. Clin J Pain 2004;20:309–18.
18. Tan G, Jensen MP, Thornby JI, Shanti BF. Validation of the Brief Pain Inventory for chronic nonmalignant pain. J Pain 2004;5:133–7.
19. Dworkin RH, Turk DC, Farrar JT, Haythornthwaite JA, Jensen MP, Katz NP, et al. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. Pain 2005;113:9–19.
20. Williams VS, Smith MY, Fehnel SE. The validity and utility of the BPI interference measures for evaluating the impact of osteoarthritic pain. J Pain Symptom Manage 2006;31:48–57.
21. Radbruch L, Loick G, Kiencke P, Lindena G, Sabatowski R, Grond S, et al. Validation of the German version of the Brief Pain Inventory. J Pain Symptom Manage 1999;18:180–7.
22. Mendoza T, Mayne T, Rublee D, Cleeland C. Reliability and validity of a modified Brief Pain Inventory short form in patients with osteoarthritis. Eur J Pain 2006;10:353–61.
23. Cleeland C. The Brief Pain Inventory: user guide. Houston: MD Anderson Cancer Center; 2009.
24. Farrar JT, Portenoy RK, Berlin JA, Kinman JL, Strom BL. Defining the clinically important difference in pain outcome measures. Pain 2000;88:287–94.
25. Farrar JT, Young JP Jr, LaMoreaux L, Werth JL, Poole RM. Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. Pain 2001;94:149–58.
26. Turk DC, Dworkin RH, Allen RR, Bellamy N, Brandenburg N, Carr DB, et al. Core outcome domains for chronic pain clinical trials: IMMPACT recommendations. Pain 2003;106:337–45.
27. Mease PJ, Spaeth M, Clauw DJ, Arnold LM, Bradley LA, Russell IJ, et al.

Estimation of minimum clinically important difference for pain in fibromyalgia. Arthritis Care Res (Hoboken) 2011;63:821–6.
28. Dworkin RH, Turk DC, Wyrwich KW, Beaton D, Cleeland CS, Farrar JT, et al. Interpreting the clinical importance of treatment outcomes in chronic pain clinical trials: IMMPACT recommendations. J Pain 2008; 9:105–21.
29. Smets EM, Garssen B, Bonke B, De Haes JC. The Multidimensional Fatigue Inventory (MFI) psychometric qualities of an instrument to assess fatigue. J Psychosom Res 1995;39:315–25.
30. Smets EM, Garssen B, Cull A, de Haes JC. Application of the multidimensional fatigue inventory (MFI-20) in cancer patients receiving radiotherapy. Br J Cancer 1996;73:241–5.
31. Barendregt PJ, Visser MR, Smets EM, Tulen JH, van den Meiracker AH, Boomsma F, et al. Fatigue in primary Sjogren's syndrome. Ann Rheum Dis 1998;57:291–5.
32. Dekkers OM, Biermasz NR, Smit JW, Groot LE, Roelfsema F, Romijn JA, et al. Quality of life in treated adult craniopharyngioma patients. Eur J Endocrinol 2006;154:483–9.
33. Jansen AJ, Essink-Bot ML, Beckers EA, Hop WC, Schipperus MR, Van Rhenen DJ. Quality of life measurement in patients with transfusion-dependent myelodysplastic syndromes. Br J Haematol 2003;121:270–4.
34. Ericsson A, Mannerkorpi K. Assessment of fatigue in patients with fibromyalgia and chronic widespread pain: reliability and validity of the Swedish version of the MFI-20. Disabil Rehabil 2007;29:1665–70.
35. Fishbain DA, Lewis J, Cole B, Cutler B, Smets E, Rosomoff H, et al. Multidisciplinary pain facility treatment outcome for pain-associated fatigue. Pain Med 2005;6:299–304.
36. Lin JM, Brimmer DJ, Maloney EM, Nyarko E, Belue R, Reeves WC. Further validation of the Multidimensional Fatigue Inventory in a US adult population sample. Popul Health Metr 2009;7:18.
37. Clauw DJ, Mease P, Palmer RH, Gendreau RM, Wang Y. Milnacipran for the treatment of fibromyalgia in adults: a 15-week, multicenter, randomized, double-blind, placebo-controlled, multiple-dose clinical trial. Clin Ther 2008;30:1988–2004.
38. Hays RD, Stewart A. Sleep measures. In: Stewart A, Ware J, editors. Measuring functioning and well-being: the medical outcomes study approach. Durham (NC): Duke University Press; 1992. p. 235–59.
39. Mease PJ, Clauw DJ, Gendreau RM, Rao SG, Kranzler J, Chen W, et al. The efficacy and safety of milnacipran for treatment of fibromyalgia. a randomized, double-blind, placebo-controlled trial. J Rheumatol 2009; 36:398–409.
40. Hays RD, Martin SA, Sesti AM, Spritzer KL. Psychometric properties of the Medical Outcomes Study Sleep measure. Sleep Med 2005;6:41–4.
41. Payne R, Mathias SD, Pasta DJ, Wanke LA, Williams R, Mahmoud R. Quality of life and cancer pain: satisfaction and side effects with transdermal fentanyl versus oral morphine. J Clin Oncol 1998;16: 1588–93.
42. Allen RP, Kosinski M, Hill-Zabala CE, Calloway MO. Psychometric evaluation and tests of validity of the Medical Outcomes Study 12-item Sleep Scale (MOS sleep). Sleep Med 2009;10:531–9.
43. Coyne KS, Zhou Z, Bhattacharyya SK, Thompson CL, Dhawan R, Versi E. The prevalence of nocturia and its effect on health-related quality of life and sleep in a community sample in the USA. BJU Int 2003;92: 948–54.
44. Wolfe F, Michaud K, Li T. Sleep disturbance in patients with rheumatoid arthritis: evaluation by medical outcomes study and visual analog sleep scales. J Rheumatol 2006;33:1942–51.
45. Unruh ML, Hartunian MG, Chapman MM, Jaber BL. Sleep quality and clinical correlates in patients on maintenance dialysis. Clin Nephrol 2003;59:280–8.
46. Rejas J, Ribera MV, Ruiz M, Masrramon X. Psychometric properties of the MOS (Medical Outcomes Study) Sleep Scale in patients with neuropathic pain. Eur J Pain 2007;11:329–40.
47. Cappelleri JC, Bushmakin AG, McDermott AM, Dukes E, Sadosky A, Petrie CD, et al. Measurement properties of the Medical Outcomes Study Sleep Scale in patients with fibromyalgia. Sleep Med 2009;10: 766–70.
48. Martin S, Chandran A, Zografos L, Zlateva G. Evaluation of the impact of fibromyalgia on patients' sleep and the content validity of two sleep scales. Health Qual Life Outcomes 2009;7:64.
49. Seidenberg M, Haltiner A, Taylor MA, Hermann BB, Wyler A. Development and validation of a Multiple Ability Self-Report Questionnaire. J Clin Exp Neuropsychol 1994;16:93–104.
50. Banos JH, LaGory J, Sawrie S, Faught E, Knowlton R, Prasad A, et al. Self-report of cognitive abilities in temporal lobe epilepsy: cognitive, psychosocial, and emotional factors. Epilepsy Behav 2004;5:575–9.
51. Martin R, Griffith HR, Sawrie S, Knowlton R, Faught E. Determining empirically based self-reported cognitive change: development of reliable change indices and standardized regression-based change norms for the multiple abilities self-report questionnaire in an epilepsy sample. Epilepsy Behav 2006;8:239–45.
52. Donovan KA, Small BJ, Andrykowski MA, Schmitt FA, Munster P, Jacobsen PB. Cognitive functioning after adjuvant chemotherapy and/or radiotherapy for early-stage breast carcinoma. Cancer 2005;104: 2499–507.
53. Jim HS, Donovan KA, Small BJ, Andrykowski MA, Munster PN, Jacobsen PB. Cognitive functioning in breast cancer survivors: a controlled comparison. Cancer 2009;115:1776–83.
54. Williams DA, Clauw DJ, Glass JM. Perceived cognitive dysfunction in fibromyalgia syndrome. J Musculoskelet Pain 2011;19:66–75.
55. Owen RT. Milnacipran hydrochloride: its efficacy, safety and tolerability profile in fibromyalgia syndrome. Drugs Today (Barc) 2008;44: 653–60.

Summary Table for Fibromyalgia Measures

Scale	Purpose/content	Method of administration	Respondent burden	Administrative burden	Score interpretation	Reliability evidence	Validity evidence	Ability to detect change	Strengths	Cautions
Fibromyalgia Impact Questionnaire	Disease-specific assessment of multiple facets of fibromyalgia: physical function, overall impact, and symptoms	Patient self-report	Completion in 3–5 minutes	Hand scored in under 5 minutes	Range: 0–10, higher scores indicate greater impairment; total score range: 0–100	Internal consistency: >0.90; test–retest: 0.56–0.95	Factor structure confirmed; construct validity supported	Minimum clinically important difference: 14% change on total score	Only disease-specific measure in fibromyalgia; useful in both clinical practice and clinical research; covers multiple aspects of fibromyalgia	Functional scale biased to high levels of disability; newer Fibromyalgia Impact Questionnaire Revised not tested in clinical trials
Brief Pain Inventory	General assessment of pain: severity interference, medications, relief from medications, presence of pain, and pain distribution	Patient self-report or interview	Completion in 5 minutes	Hand scored by summing in under 5 minutes	Pain severity (range 0–10) with higher scores indicating greater pain, interference (range 0–10) with higher scores indicating greater interference	Internal consistency: 0.85–0.88 for severity and interference, respectively; test–retest: 0.83–0.98	Factor structure confirmed; construct validity supported	Minimum clinically important difference: 30% change on severity score	Rigorous development; widely used clinically; widely used in clinical trials, recommended by Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials group	Scores other than severity and interference often not reported
Multidimensional Fatigue Inventory	Profound fatigue associated with illness: general, physical, mental, reduced motivation, and reduced activity	Patient self-report	Completion in 5 minutes	Hand scored in 5 minutes	Each scale contains a minimum value of 4 and a maximum of 20; higher scores indicate greater fatigue	Internal consistency: average is 0.80 for scales; total score internal consistency = 0.93	Factor structure confirmed; construct validity supported	No minimum clinically important difference	Used in fibromyalgia clinical trials; good metric properties for clinical and research use	Response set can be confusing to respondents
Medical Outcomes Study Sleep Scale	A generic assessment of sleep problems: disturbance, duration, adequacy, somnolence, snoring, shortness of breath, and summary indices	Patient self-report	Completion in 3–5 minutes	Hand scored in 5–7 minutes	Most scales range 0–100, scales are mixed with regard to the interpretation of scale values	Internal consistency range 0.64–0.84; in fibromyalgia, internal consistency = 0.70	Factor structure confirmed; construct validity supported	Minimal important difference 5.1	Generic instrument; widely used in clinical research; responsive to change; can be used in clinical settings	Scoring and meaning are not intuitive
Short Form 36 physical and mental component scores	Physical and mental health functional status	Self-report, interview, or online	Completion in 5–7 minutes	Scored by publisher, software online, or distributed	Lower scores indicate worse functional status; scores are norm-based on US population with a mean of 50; a 10-point change is equivalent to 1 SD	Internal consistency >0.90; test– retest: r ranges 0.60–0.80 for 2 weeks	Factor structure confirmed; construct validity supported	Minimum clinically important difference in fibromyalgia: 6-point change	Extremely well-developed instrument; useful in clinical work; useful in clinical trials	Can be costly to use; scoring is complex and requires software
Multiple Ability Self-Report Questionnaire	Perceived cognitive deficits: language, visual-perceptual, verbal memory, visual-spatial memory, and attention-concentration	Self-report	Completion in 10 minutes	Hand scored involving item reversal and summing	Higher scores indicate more dysfunction; 8-item scales have a max value of 40; 6-item scales have max value of 30; total score has max value of 190	Internal consistency >0.70, test–retest: 0.72–0.79	Construct validity supported using convergent and divergent methods	Minimum clinically important differences have not been established for fibromyalgia	Only instrument that assesses multiple aspects of dyscognition in fibromyalgia, useful clinically; useful for clinical trials of fibromyalgia	Somewhat lengthy
Hospital Anxiety and Depression Scale	Anxiety and depression screener for nonpsychiatric populations	Self-report	Completion in 2–5 minutes	Hand scoring by summing items (1–2 minutes)	Higher scores indicate more anxiety and depressive symptoms; 0–7 normal, 8–10 mild, >11 probable caseness	Internal consistency 0.80–0.93	Factor structure confirmed; construct validity supported	Minimum clinically important difference not established for fibromyalgia	Avoids somatic symptoms and extreme psychiatric symptoms more appropriate for a nonpsychiatric population; quick to administer and score in the clinic or research setting; has been used in multiple fibromyalgia clinical trials	Does not provide a diagnosis, only an estimate of potential caseness