Research report

Development and validation of a patient-report measure of fatigue associated with depression

Louis S. Matza a,⁎, Glenn A. Phillips b, Dennis A. Revicki a, Lindsey Murray a, Karen G. Malley c

a Center for Health Outcomes Research, United BioSource Corporation, Bethesda, MD, United States
b Formerly with Eli Lilly & Company, Indianapolis, IN, United States
c Malley Research Programming, Inc., Rockville, MD, United States

1. Introduction

Fatigue is one of the most commonly occurring symptoms of major depressive disorder (Marcus et al., 2005; Maurice-Tison et al., 1998; Tylee et al., 1999) and it is associated with substantial impairment in social and work functioning (Swindle et al., 2001). Fatigue is also a frequent residual symptom that may persist in roughly 20% to 38% of patients who have remitted following treatment (Barkham et al., 1996; Nierenberg et al., 1999, 2010). Residual symptoms are an important target for research and treatment because they are predictive of relapse of depressive episodes (Judd et al., 1998, 2000; Menza et al., 2003; Mintz et al., 1992; Thase et al., 1992). Furthermore, residual symptoms continue to contribute to
psychosocial and occupational impairment even after other symptoms of depression have resolved with pharmacological or psychological treatment (Judd et al., 2000; Kennedy and Paykel, 2004; Nierenberg et al., 2010; Ogrodniczuk et al., 2004; Paykel, 1998). Consequently, there is growing interest in the effectiveness of antidepressants for treatment of specific residual symptoms of depression, such as fatigue (Arnold, 2008; Demyttenaere et al., 2005; Menza et al., 2003).

Despite the prevalence of fatigue and its importance as a residual symptom, there is no validated patient-reported outcome (PRO) measure designed specifically to assess fatigue and its impact among patients with depression (Ferentinos et al., 2007). Commonly used depression-specific measures do not provide detailed assessment of fatigue. For example, the Hamilton Depression Rating Scale (HAM-D) has two items with response options that mention fatigue or loss of energy, but no items specifically assessing fatigue (Hamilton, 1960). The Montgomery–Asberg Depression Rating Scale (MADRS) includes only one related item, which focuses on lassitude (Montgomery and Asberg, 1979). More recently developed instruments such as the Inventory for Depressive Symptomatology (IDS) and Chicago Multiscale Depression Inventory (CMDI) include at least one item assessing fatigue (Nyenhuis et al., 1998; Rush et al., 1986), but they do not provide a thorough or multidimensional assessment of fatigue, and they do not assess the impact of fatigue. The only depression-specific instrument providing a more detailed perspective is the Motivation and Energy Inventory (Fehnel et al., 2004). However, this instrument is not ideal for assessment of “fatigue” because it focuses on “motivation and energy,” which are related, but different constructs.

Generic instruments (i.e., measures designed to be completed by respondents regardless of medical or psychiatric condition) are available for a more detailed assessment of fatigue. However, none of these instruments, such as the Brief Fatigue Inventory (Mendoza et al., 1999) and the Fatigue Assessment Scale (Michielsen et al., 2003), were developed specifically for patients with depression. One available generic questionnaire, the Fatigue Questionnaire (FQ), has demonstrated reliability and validity in a sample of 81 patients with depression, but content validity was not examined (Ferentinos et al., 2010). Furthermore, five of the FQ’s 11 items assess cognition (e.g. memory problems, slips of the tongue) rather than fatigue, suggesting that the content of this questionnaire may not be focused on important aspects of fatigue associated with depression. To be considered acceptable for assessment of treatment outcomes, PRO instruments must have demonstrated content validity and good measurement properties in the specific target population (Food and Drug Administration, FDA, 2009; Leidy et al., 1999). The generic fatigue measures do not appear to meet these standards for use in patients with depression.

In sum, no available instrument was located that would be considered adequate for a detailed assessment of fatigue and its impact in clinical trials of treatments for depression. Thus, the purpose of the current study was to develop and validate a PRO measure designed to assess depression-related fatigue and its impact on patients’ lives. This measure, called the Fatigue Associated with Depression Questionnaire (FAuD) was developed in accordance with recommendations in the Food and Drug Administration PRO Guidance Document (Food and Drug Administration, FDA, 2009). The FAuD was initially drafted and refined based on literature review and qualitative research with clinicians and patients diagnosed with depression. Then, a draft version of the questionnaire was administered in a validation study conducted with a sample of patients with depression. Statistical analysis first focused on item reduction and subscale identification, followed by assessment of the reliability and validity of the final version of the FAuD.

2. Qualitative instrument development

2.1. Patient focus groups

As the first step in gathering information to generate items for the FAuD, four focus groups were conducted at two clinical sites (northwest and southeast United States) to identify key aspects of fatigue and its impact, while identifying language used by patients to describe fatigue. The sample of 20 patients diagnosed with depression was 60% female, 85% white, with a mean age of 46.4 years, and 16 were currently being treated with antidepressant medication. Words and phrases used by participants to describe their experience of fatigue included “fatigue,” “tired,” “effort,” “slowed down,” lack of “motivation,” lack of “energy,” “exhausted,” “wiped out,” “worn out,” “drained,” “heavy,” “weak,” and “paralyzing.” Participants also provided detailed descriptions of the impact of fatigue on multiple domains of their lives, including work/productivity, social activities, relationships with significant others (including sexual activity), and activities of daily living.

Although larger sample sizes can contribute to confidence in content validity, saturation is the point at which additional sampling offers no new information (Leidy and Vernon, 2008). It was determined upon sequential review of focus group transcripts that saturation was reached after the second focus group. Qualitative data gathered during the focus groups provided a rich source of information and language that was used when drafting the FAuD, ensuring that the items of the FAuD were directly linked to patient perceptions and descriptions of fatigue.

2.2. Clinician input

Following the patient focus groups, 14 clinicians (psychiatrists and psychologists) were interviewed to assess their impressions of patients’ experience with fatigue associated with depression. All respondents stated that fatigue was a common symptom with broad functional impact among patients with depression. The clinicians provided additional descriptors and examples of how patients describe their fatigue and its impact on their lives.

2.3. Cognitive debriefing interviews

The FAuD was drafted based on the literature review, patient focus groups, and clinician interviews. The preliminary FAuD was then administered to patients with depression in a cognitive debriefing interview study (n = 18; 72.2% female; 72.2% white; mean age = 37.1 years; 66.7% currently treated with antidepressants). The purpose of these interviews was to confirm the content validity of the FAuD.
items, while evaluating the instrument in terms of ease-of-use, clarity, comprehensibility, potential errors of omission, and possible redundancy. All participants reported that the preliminary FAsD was clear and relevant to their condition. Suggestions for changes to the questionnaire were minor. An item asking about “weakness” was the only item that was unclear to a substantial number of participants (n=7). However, because other participants said this item was relevant, it was retained and clarified as “physically weak.”

Response options for the FAsD are presented as a five-point Likert-type frequency scale because most participants talked about fatigue symptoms in terms of frequency rather than severity (e.g., how often they felt tired). Fatigue impact items were worded in terms of severity for clarity and ease of reading (i.e., how severe was the impact). In the cognitive interviewing study, there were no concerns reported by the participants regarding these response options. A recall period for the FAsD of 1 week was selected to limit a recall bias that may occur with longer recall periods while still providing a sufficient duration to capture the experience and impact of fatigue. While a few participants in cognitive debriefing interviews indicated that they would prefer a longer recall period than 1 week, the majority indicated that a one week period was appropriate. Therefore, the one week recall period was retained. Minor edits were made to the questionnaire based on participants’ comments, yielding a 16-item draft of the FAsD to be administered in the subsequent psychometric validation study.

3. Methods: instrument validation

3.1. Study design and sample selection

A psychometric validation study was conducted to refine the draft 16-item FAsD and evaluate the reliability and validity of the final instrument. Data were collected from patients with depression at 13 clinical sites in the United States (including 10 psychiatric sites, one primary care site, and one site providing both psychiatric and primary care services). Inclusion criteria were: age ≥18, clinical diagnosis of depression, and current symptoms of depression as indicated by a score of ≥5 on the 8-item Patient Health Questionnaire (PHQ-8) (Kroenke and Spitzer, 2002). Exclusion criteria were diagnosis of bipolar disorder; receiving treatment with a mood stabilizer or antipsychotic; or diagnosed with the following medical conditions that could cause fatigue: chronic fatigue syndrome, sleep apnea, cancer, multiple sclerosis, or HIV. The validation study and the cognitive debriefing interviews were approved by an Institutional Review Board, and all participants provided written informed consent.

Two groups of patients were recruited: “maintenance patients” and “new treatment patients.” Maintenance patients had not begun a new treatment for depression within the past 14 days, expected no change in treatment of depression in the next 2 weeks, and may or may not have been receiving treatment for depression. “New treatment patients” began treatment with a new antidepressant within the past 7 days because of clinically indicated reasons.

To assess test–retest reliability, participants’ symptoms and treatment should remain stable between questionnaire administrations to ensure that score changes primarily represent measurement error rather than true change in patients’ condition (Leidy et al., 1999). Therefore, a measure of symptom stability (i.e., the Overall Treatment Effect Scale) was administered to the maintenance patients at a second study visit 14±4 days after the first visit, and test–retest reliability of the FAsD was assessed in the stable subgroup.

3.2. Measures

3.2.1. Fatigue Associated with Depression Questionnaire (FAsD)

The 16-item draft of the FAsD administered in this study included seven questions about fatigue symptoms with a response scale of “never,” “rarely,” “sometimes,” “often,” and “always,” in addition to nine questions about the impact of fatigue with a response scale of “not at all,” “a little,” “somewhat,” “quite a bit,” and “very much.”

3.2.2. Brief Fatigue Inventory (BFI)

The Brief Fatigue Inventory includes three items assessing the severity of fatigue and six items assessing the degree to which fatigue has interfered with a range of domains, including mood, walking ability, and enjoyment of life (Mendoza et al., 1999). Scores range from 0 to 10, with higher scores representing greater fatigue.

3.2.3. Inventory of Depressive Symptomatology-Self-Report (IDS-SR)

The 30-item self-report version of the Inventory of Depressive Symptomatology (IDS-SR) was designed to assess the severity of depressive symptoms, including all DSM-IV depression diagnostic criteria as well as commonly associated symptoms (Rush et al., 1986, 1996). Scores may range from 0 to 84, with higher scores representing greater severity of depression.

3.2.4. Epworth Sleepiness Scale (ESS)

The Epworth Sleepiness Scale (ESS) (Johns, 1991) is used to determine the level of daytime sleepiness. Patients rate the chance of dozing or sleeping during eight activities. Scores may range from 0 to 24, with higher scores representing greater sleepiness.

3.2.5. SF-36 Version 1.0

The SF-36 is a generic health-related quality of life questionnaire that yields a Mental Component Summary (MCS-12) and a Physical Component Summary (PCS-12) as well as eight domain scores (Ware and Sherbourne, 1992). Summary and subscale scores range from 0 to 100 with higher scores indicating better functioning.

3.2.6. Clinical Global Impression-Severity (CGI-S) Scale

The Clinical Global Impression-Severity Scale (Guy, 1976) was completed by clinicians to assess their overall impression of the severity of the patient’s depressive illness. The score ranges from 1 (normal, not at all ill) to 7 (among the most extremely ill patients).

3.2.7. Overall Treatment Effect Scale

The maintenance patients reported their change in fatigue from their first visit to the second visit 14±4 days later using
the Overall Treatment Effect Scale (OTE) (Jaeschke et al., 1989; Juniper et al., 1994). The first question asked participants to indicate whether their fatigue associated with depression had improved, remained the same, or worsened. If participants indicated that their fatigue had changed, they rated the degree of change on a 7-point scale. Improvement ranged from (1) “Almost the same, hardly better at all” to (7) “A very great deal better,” and worsening ranged from (−1) “Almost the same, hardly worse at all” to (−7) “A very great deal worse” (−7). Participants with scores of −1, 0, or 1 considered to be stable between the two visits, and their data were used in test–retest reliability analyses.

3.2.8. Demographic and clinical forms
All participants completed a brief demographic and clinical form. Clinicians completed a clinical information form for each participant, reporting diagnoses, severity of depression, comorbid conditions, and medications.

3.3. Statistical analysis
Initial analyses focused on item reduction and subscale identification to determine the final version of the FasD, followed by subsequent analyses to examine the reliability and validity of the final FasD. Analyses for item reduction, subscale identification, internal consistency reliability, and validity were conducted using data pooled from both maintenance patients and new treatment patients, while test–retest reliability was assessed only in the stable subgroup of maintenance patients.

SAS statistical software version 8.2 was used for all analyses. Categorical variables were summarized in terms of frequencies and percentages. For each continuous variable, the mean, standard deviation, median, range, percent at floor, and percent at ceiling are presented.

The final three items of the 16-item draft FasD were not completed by all participants because Item 14 was applicable only to participants with a significant other, and Items 15 and 16 were applicable only to respondents currently attending work or school. Exploratory factor analyses and analysis of internal consistency reliability were conducted with the subset of participants who responded to all relevant items.

3.3.1. Item reduction and subscale identification
Several factors were considered when performing item reduction. Items with the following characteristics were considered for removal: ceiling or floor effect with more than 50% of respondents choosing the highest or lowest response options, redundancy of measurement as demonstrated by a correlation with another item (r ≥ 0.85), and greater than 5% missing responses.

Item reduction was also influenced by results of an exploratory factor analysis (EFA) with oblique rotation (Promax rotation method: assumes correlated factors), using principal axis factoring extraction. The number of factors was set a priori to no factors. Eigen values down to 0.7 were used to examine potential factor structures (Stull et al., 2007). Items were considered for deletion if they did not load on any factor ≥ 0.4 or loaded on more than one factor ≥ 0.4. When performing item reduction, the clinical importance and qualitative results of patient focus groups and interviews were also considered. After items were deleted, the EFA was conducted again to ascertain if any other items met the removal criteria.

3.3.2. Psychometric evaluation of the final FasD
Internal consistency reliability of the FasD was assessed for each subscale and the total score using Cronbach’s formula for coefficient alpha. Cronbach’s alpha values greater than 0.70 but less that 0.95 are generally considered to be acceptable (Hays et al., 1998; Nunnally and Bernstein, 1994). Test–retest reliability of the FasD was assessed in the stable subsample of maintenance patients, as indicated by OTE responses (i.e., stable = OTE score of −1, 0, or 1). Intraclass correlation coefficients (ICC) were conducted to evaluate the degree of association between FasD scores at Visit 1 and Visit 2. An ICC ≥ 0.6 is generally thought to indicate acceptable test–retest reliability (Landis and Koch, 1977).

Two types of construct validity were examined with Spearman correlations of the FasD with previously validated questionnaires (BFI, IDS, ESS, SF-36). Convergent validity was supported when the FasD is substantially correlated with subscales measuring similar concepts. Conversely, divergent validity was supported when correlations with scales measuring different concepts, such as sleepiness, were relatively weak.

Known-groups validity was assessed by categorizing participants into depression severity groups as determined by CGI-S ratings, and examining the pattern of FasD scores among these subgroups.

4. Results: instrument validation
4.1. Sample description
A total of 320 participants were enrolled, but three were excluded from analyses due to violations of inclusion/exclusion criteria. Table 1 summarizes sample characteristics of a per protocol analysis sample (N = 317), as well as the subgroups of maintenance patients and new treatment patients. The per protocol sample was primarily female (68.1%), with a mean age of 47.0 years. The majority of the sample was white (60.3%), not married (64.7%), and employed full-time (38.5%) or part-time (15.5%). The mean age at the time of the first depression diagnosis was 34.6 years, and 86.1% of the sample was receiving medication for the treatment of depression at the time of study enrollment (including maintenance and new antidepressant treatments). The most common classes of antidepressant treatment were selective serotonin reuptake inhibitors (46.7%) and serotonin–norepinephrine reuptake inhibitors (30.9%). The majority of participants reported no comorbid medical conditions (52.4%). The most common comorbid medical conditions were hypertension (21.5%) and arthritis (18.5%). Most participants reported no current or previous psychiatric diagnoses (69.4%) other than depression. The most common comorbid psychiatric conditions were anxiety disorders (23.0%). Compared with the maintenance subgroup, the new treatment subgroup was significantly younger with a higher percentage of male participants. The new treatment subgroup was also less likely to be white and less likely to be married.
4.2. Item analysis, item reduction, and subscale identification

The descriptive statistics, including mean scores and rates of missing data, for each individual item of the draft FAsD are presented in Table 2. Mean item scores range from 2.0 (Item 15: Prevented you from going to work or school) to 3.9 (Item 2: Tired). Among the first 13 items designed to be answered by all respondents, there was minimal missing data. Similarly, among participants who reported having a spouse/significant other, there was minimal missing data on Item 14. Among participants reported having a job or going to school, there was minimal missing data on Items 15 or 16. Item 15 (Prevented you from going to work or school) had a floor effect, with 51.4% of participants choosing the lowest response option, suggesting that this item was not relevant to the majority of participants. In addition, this item was conceptually redundant with Item 16 (Limited your productivity at work or school), and consequently, it was deleted prior to conducting the EFA. No item had ceiling effects for more than 27% of the sample.

The EFA was conducted with the remaining 15 items (Items 1–14 and 16), and the top three eigenvalues were 7.85, 0.95, and 0.52. Because it was specified a priori that factors with eigenvalues greater than 0.70 would be retained, two factors were retained. Six items clearly loaded on factor 1, with loadings greater than 0.55 on factor 1 and less than 0.31 on factor 2. Seven items clearly loaded on factor 2, with factor loadings greater than 0.46 on factor 2 and less than 0.37 on factor 1. Two items did not fit clearly within this two-factor solution. Item 6 cross-loaded, with factor loadings greater than 0.40 on both factors. Item 13 did not appear to load on either factor (factor loadings of 0.36 on factor 1 and 0.27 and factor 2). These two items were deleted, resulting in the final 13-item instrument.

4.3. Factor structure of the final 13-item FAsD

After dropping the three items as described above, the EFA was re-run with the final 13 items. There were two eigenvalues greater than 0.70 (top three eigenvalues were 6.88, 0.94, and 0.47). Therefore, two factors were retained, and factor loadings are presented in Table 3. The two factors were consistent with the hypothesized domains of “fatigue experience” and “fatigue impact.” Items 1, 2, 3, 4, 5, and 7 loaded on the experience factor, with all loadings greater than 0.53. Items 8, 9, 10, 11, 12, 14, 16 loaded on the impact factor, with all loadings greater than 0.47. All subsequent analyses were conducted with this final 13-item version of the FAsD, including a 6-item fatigue experience subscale, a 7-item fatigue impact subscale, and a total score (see Appendix A for the 13-item FAsD).

4.4. Descriptive statistics and scoring of the 13-item FAsD

Three scores are computed for the final 13-item FAsD: experience subscale (items 1 to 6), impact subscale (7 to 13), and a total score (1–13). All three scores are computed as the mean of answered items within the scale. There are likely to be respondents without responses to items 12 (intimate relationships) and/or 13 (work or school) because these two items are not relevant to all respondents. Therefore, when respondents intentionally leave one or both of these items blank because they do not currently have a job or a relationship, the items should be considered “not applicable” rather than missing.

In the current study, missing data were unlikely to affect the analyses because only six participants were missing a response to one item, and no participants were missing more than one response. In future studies, it is recommended that the experience subscale should not be computed if there are

---

**Table 1**

Demographic and clinical characteristics.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Total sample (N=317)</th>
<th>Maintenance patients (N=199)</th>
<th>New patients (N=118)</th>
<th>p-value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender (n, %)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>101 (31.9%)</td>
<td>54 (27.1%)</td>
<td>47 (39.8%)</td>
<td>0.019</td>
</tr>
<tr>
<td>Female</td>
<td>216 (68.1%)</td>
<td>145 (72.9%)</td>
<td>71 (60.2%)</td>
<td></td>
</tr>
<tr>
<td>Age (mean, SD)</td>
<td>47.0, 13.0</td>
<td>49.7, 13.1</td>
<td>42.5, 11.6</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Racial background (n, %)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asian or Pacific Islander</td>
<td>17 (5.4%)</td>
<td>6 (3.0%)</td>
<td>11 (9.3%)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Black, not of Hispanic origin</td>
<td>44 (13.9%)</td>
<td>8 (4.0%)</td>
<td>36 (30.5%)</td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>49 (15.5%)</td>
<td>32 (16.1%)</td>
<td>17 (14.4%)</td>
<td></td>
</tr>
<tr>
<td>White, not of Hispanic origin</td>
<td>191 (60.3%)</td>
<td>143 (71.9%)</td>
<td>48 (40.7%)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>16 (5.0%)</td>
<td>10 (5.0%)</td>
<td>6 (5.1%)</td>
<td></td>
</tr>
<tr>
<td>Marital status (n, %)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>112 (35.3%)</td>
<td>79 (39.7%)</td>
<td>33 (28.0%)</td>
<td>0.035</td>
</tr>
<tr>
<td>Not married</td>
<td>205 (64.7%)</td>
<td>120 (60.3%)</td>
<td>85 (72.0%)</td>
<td></td>
</tr>
<tr>
<td>Employment status (n, %)</td>
<td></td>
<td></td>
<td></td>
<td>0.27</td>
</tr>
<tr>
<td>Full-time work</td>
<td>122 (38.5%)</td>
<td>80 (40.2%)</td>
<td>42 (35.6%)</td>
<td></td>
</tr>
<tr>
<td>Part-time work</td>
<td>49 (15.5%)</td>
<td>34 (17.1%)</td>
<td>15 (12.7%)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>146 (46.1%)</td>
<td>85 (42.7%)</td>
<td>61 (51.7%)</td>
<td></td>
</tr>
<tr>
<td>Age at time of first depression diagnosis (mean, SD)</td>
<td>34.6, 13.5</td>
<td>35.5, 14.0</td>
<td>33.1, 12.6</td>
<td>0.13</td>
</tr>
<tr>
<td>Currently receiving medication for the treatment of depression (n, %)</td>
<td>273 (86.1%)</td>
<td>155 (77.9%)</td>
<td>118 (100.0%)</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

* P-value is for comparisons between the maintenance patients and new treatment patients. Continuous variables compared with t-tests; categorical variables compared with chi-square analyses.
more than two missing responses, the impact subscale should not be computed if there are more than two missing responses to applicable items (i.e., items 12 and 13 are not applicable to all respondents), and a total score should not be computed with more than four missing responses to applicable items.

Descriptive statistics for the FAsD subscales are presented in Table 2. Mean scores were 3.51 (experience subscale), 3.17 (impact subscale), and 3.34 (total score). The range for both subscales and the total score was 1.0 to 5.0. There were no substantial floor or ceiling effects, and there were no missing data for subscale or total scores.

4.5. Reliability

The FAsD demonstrated adequate internal consistency reliability, with Cronbach’s alphas of 0.90 for the experience subscale, 0.88 for the impact subscale, and 0.93 for the total score (Table 4). Eliminating any individual item did not substantially increase the alpha for either subscale or the total score.
of FAsD scores across subgroups of participants differing by the SF-36 social functioning scale (stronger correlation than the FAsD experience subscale with the SF-36 vitality scale, which assesses the “experience” of feeling energetic, was stronger with the FAsD experience subscale than with the FAsD impact subscale (−0.68 versus −0.53). In contrast, the FAsD impact subscale had a stronger correlation with the SF-36 social functioning scale (−0.70 versus −0.55), which focuses on impact rather than experience. In sum, these correlational analyses demonstrated convergent and divergent validity of the FAsD total score, while supporting the distinction between the experience and impact subscales.

The strong correlations of the FAsD scales with the BFI suggest that these two questionnaires capture similar aspects of fatigue. However, there are two key differences between the questionnaires. First, the BFI was designed primarily for cancer patients, with a general structure derived from the brief pain inventory (Mendoza et al., 1999). In contrast, the FAsD is the first fatigue measure that was developed based on the perceptions and language of patients with depression, followed by a psychometric evaluation performed in the target population as is frequently recommended for PRO measures (Food and Drug Administration, FDA, 2009; Leidy et al., 1999). Second, whereas the BFI yields only a global score supported by strong signal-factor model fit in its initial validation study (Mendoza et al., 1999), the FAsD allows for separate evaluation of experience and impact, with two validated subscales. In focus groups conducted for the current study, patients with depression

5. Discussion

The FAsD appears to be a useful measure of fatigue experience and impact among patients with depression. In this initial psychometric evaluation, item reduction resulted in the final 13-item version of the FAsD, with a 6-item experience subscale and a 7-item impact subscale. Measurement properties of the questionnaire were strongly supported, with findings suggesting good internal consistency reliability, test–retest reliability, convergent validity, divergent validity, and known-groups validity. Furthermore, minimal missing data in the validation study and patients’ feedback during cognitive debriefing interviews support the instrument’s ease-of-use and comprehensibility.

In the correlation analyses examining construct validity, logical patterns emerged (Table 5). The FAsD total score was most strongly related to the generic measure of fatigue (BFI) and the depression symptom measure (IDS). While still statistically significant, the correlation of the FAsD total score with a sleep measure (ESS) was of a substantially weaker magnitude, suggesting that fatigue and sleepiness are related, but distinct constructs. Results of correlations involving the FAsD subscales generally followed similar patterns as those for the total score, although correlations of the FAsD subscales with the SF-36 suggest that patients generally understood the distinction between fatigue experience and impact. For example, the magnitude of the correlation with the SF-36 vitality subscale, which assesses the “experience” of feeling energetic, was stronger with the FAsD experience subscale than with the FAsD impact subscale (−0.68 versus −0.53). In contrast, the FAsD impact subscale had a stronger correlation with the SF-36 social functioning scale (−0.70 versus −0.55), which focuses on impact rather than experience. In sum, these correlational analyses demonstrated convergent and divergent validity of the FAsD total score, while supporting the distinction between the experience and impact subscales.

4.6. Validity

All correlations of the FAsD subscales and total score with previously validated questionnaires were statistically significant (p < 0.0001; Table 5). The total score demonstrated convergent validity, as indicated by strong correlations with other measures assessing fatigue/energy including the BFI (r = 0.84) and the SF-36 vitality scale (r = −0.65). The FAsD was also highly correlated with depression symptoms as assessed by the IDS (r = 0.71). Divergent validity of the FAsD was supported by the weaker correlation with the ESS (r = 0.37), suggesting that fatigue and sleepiness are related, yet distinct.

The results for the FAsD subscales followed similar patterns to the correlations for the total score, although some notable differences between the two subscales were apparent (Table 5). For example, the FAsD experience subscale was more strongly correlated with the SF-36 vitality scale (−0.68 vs. −0.53). In contrast, the FAsD impact subscale had a stronger correlation than the FAsD experience subscale with the SF-36 social functioning scale (−0.70 vs. −0.55).

Known-groups validity was demonstrated by the pattern of FAsD scores across subgroups of participants differing by CGI-S ratings (Table 6). Higher scores on the FAsD scales were consistently associated with more severe clinician ratings of depression on the CGI-S, with statistically significant differences between severity groups. For example, the mean FAsD total score was 4.0 for patients rated by clinicians as “markedly or severely ill,” 3.4 for patients rated as “moderately ill,” 2.9 for patients rated as “mildly ill,” 2.8 for patients rated as “borderline ill,” and 2.3 for patients rated as “normal, not at all ill.” Results for the FAsD experience and impact subscales followed the same pattern as the total score.

Table 4

Internal consistency and test–retest reliability of the FAsD.

<table>
<thead>
<tr>
<th>FAsD scales</th>
<th>Internal consistency</th>
<th>Test–retest reliability: ICC (n = 122)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>Cronbach’s alpha</td>
<td></td>
</tr>
<tr>
<td>314</td>
<td>0.90</td>
<td>0.78</td>
</tr>
<tr>
<td>120</td>
<td>0.88</td>
<td>0.84</td>
</tr>
<tr>
<td>120</td>
<td>0.93</td>
<td>0.85</td>
</tr>
</tbody>
</table>

ICC = intraclass correlation coefficient.
consistently reported that fatigue has a powerful impact on multiple aspects of their lives. The FAsD provides a way to quantify this impact. Therefore, despite the considerable overlap between the FAsD and BFI, the FAsD is likely to have advantages for studies in depression.

It should be noted that the FAsD does not include items assessing impact of fatigue on cognition or thinking. Some previous fatigue instruments assess cognitive impact and distinguish between mental and physical fatigue (Chalder et al., 1993; Fehnel et al., 2004; Fisk et al., 1994; Yang and Wu, 2005), while others do not (Cella, 1997; Krupp et al., 1989; Mendoza et al., 1999; Tack, 1990). As a result of input from clinicians and patients during the qualitative phase of instrument development, FAsD does not include items assessing cognitive impact. The 14 clinicians who were interviewed were asked about the distinction between mental and physical fatigue. Eight of the 14 expressed the opinion that it was not a clear distinction, and only five of the 14 expressed the opinion that patients could distinguish between mental and physical fatigue. Furthermore, the 20 patients in focus groups never mentioned difficulty with concentration as a symptom of depression, they did not indicate that they perceived cognitive symptoms to be related to fatigue. Based on these qualitative findings, it was decided that the FAsD would focus specifically on fatigue without integrating cognitive outcomes.

One limitation of the FAsD validation study stems from the wide range of current treatments received by the sample. Most of the participants (86.1%) were receiving antidepressant treatment at the time they completed the study measures. Given that some antidepressants have the potential to cause fatigue, it is not possible to know the extent to which scores in the FAsD were influenced by patients’ current treatment rather than the depression itself. A comparison sample of patients not receiving pharmacological treatment would be required to examine this question.

PRO measure validation is an ongoing process (Revicki et al., 2000), and confidence in a new questionnaire develops based on gradually accumulating psychometric data. Therefore, despite the encouraging results, the current study should be considered only the first step in validating the FAsD. The next steps will be to examine the instrument’s responsiveness to change over time and to derive interpretation guidelines such as the minimally important difference (Revicki et al., 2008).

In sum, the FAsD demonstrated reliability and validity as a PRO measure that allows for a brief yet detailed assessment of wide range of current treatments received by the sample. The FAsD scales Normal, not at all ill (N=9), Borderline ill (N=21), Mildly ill (N=51), Moderately ill (N=154), Markedly or severely ill (N=41).

### Table 5
Construct validity: Spearman correlations\(^a\) of the FAsD with other patient reported measures.

<table>
<thead>
<tr>
<th>Participant-reported measures</th>
<th>N</th>
<th>FAsD experience subscale score</th>
<th>FAsD impact subscale score</th>
<th>FAsD total score</th>
</tr>
</thead>
<tbody>
<tr>
<td>BFI Score</td>
<td>316</td>
<td>0.757</td>
<td>0.793</td>
<td>0.835</td>
</tr>
<tr>
<td>IDS Score</td>
<td>317</td>
<td>0.619</td>
<td>0.699</td>
<td>0.714</td>
</tr>
<tr>
<td>ESS Score</td>
<td>311</td>
<td>0.352</td>
<td>0.334</td>
<td>0.367</td>
</tr>
<tr>
<td>SF-36 physical functioning</td>
<td>317</td>
<td>−0.459</td>
<td>−0.431</td>
<td>−0.450</td>
</tr>
<tr>
<td>SF-36 role limitations due to physical health</td>
<td>317</td>
<td>−0.402</td>
<td>−0.484</td>
<td>−0.486</td>
</tr>
<tr>
<td>SF-36 role limitations due to emotional problems</td>
<td>316</td>
<td>−0.386</td>
<td>−0.465</td>
<td>−0.463</td>
</tr>
<tr>
<td>SF-36 vitality</td>
<td>317</td>
<td>−0.677</td>
<td>−0.533</td>
<td>−0.646</td>
</tr>
<tr>
<td>SF-36 mental health</td>
<td>317</td>
<td>−0.433</td>
<td>−0.490</td>
<td>−0.498</td>
</tr>
<tr>
<td>SF-36 social functioning</td>
<td>317</td>
<td>−0.547</td>
<td>−0.700</td>
<td>−0.675</td>
</tr>
<tr>
<td>SF-36 pain</td>
<td>316</td>
<td>−0.439</td>
<td>−0.442</td>
<td>−0.476</td>
</tr>
<tr>
<td>SF-36 general health</td>
<td>317</td>
<td>−0.430</td>
<td>−0.385</td>
<td>−0.438</td>
</tr>
<tr>
<td>SF-36 physical component summary score</td>
<td>315</td>
<td>−0.427</td>
<td>−0.400</td>
<td>−0.450</td>
</tr>
<tr>
<td>SF-36 mental component summary score</td>
<td>315</td>
<td>−0.456</td>
<td>−0.526</td>
<td>−0.527</td>
</tr>
</tbody>
</table>

\(^a\) All are significant at \(p<0.0001\).

### Table 6
Known-groups validity of the FAsD: ANOVA\(^a\) by CGI-S ratings.

<table>
<thead>
<tr>
<th>FAsD scales</th>
<th>Normal, not at all ill (N=9)</th>
<th>Borderline ill (N=21)</th>
<th>Mildly ill (N=51)</th>
<th>Moderately ill (N=154)</th>
<th>Markedly or severely ill (N=41)</th>
<th>Overall F value</th>
<th>Significant pairwise comparisons</th>
</tr>
</thead>
<tbody>
<tr>
<td>FAsD total score</td>
<td>2.296 (0.252)</td>
<td>2.766 (0.165)</td>
<td>2.948 (0.106)</td>
<td>3.437 (0.061)</td>
<td>3.995 (0.118)</td>
<td>19.06***</td>
<td>C***, D***, F**, G***, H**, J**</td>
</tr>
<tr>
<td>FAsD experience subscale score</td>
<td>2.630 (0.258)</td>
<td>3.032 (0.169)</td>
<td>3.229 (0.108)</td>
<td>3.584 (0.062)</td>
<td>4.045 (0.121)</td>
<td>11.84***</td>
<td>C*, D***, G***, I***, J*</td>
</tr>
<tr>
<td>FAsD impact subscale score</td>
<td>1.979 (0.295)</td>
<td>2.526 (0.193)</td>
<td>2.681 (0.124)</td>
<td>3.293 (0.071)</td>
<td>3.937 (0.138)</td>
<td>19.21***</td>
<td>C***, D***, F**, G***, H**, J**</td>
</tr>
</tbody>
</table>

Pairwise comparisons: A: Normal not at all ill vs. Borderline ill; B: Normal not at all ill vs. Mildly ill; C: Normal not at all ill vs. Moderately ill; D: Normal not at all ill vs. Markedly or Severely ill; E: Borderline ill vs. Mildly ill; F: Borderline ill vs. Moderately ill; G: Borderline ill vs. Markedly or Severely ill; H: Mildly ill vs. Moderately ill; I: Mildly ill vs. Markedly or Severely ill; J: Moderately ill vs. Markedly or Severely ill.

\(^a\) ANOVA with Scheffe’s post-hoc pairwise comparisons.

\(^b\) This group includes 35 patients rated as markedly ill and six patients rated as severely ill. The two groups were combined because of the small number of patients in the severely ill group.
has started to focus on a symptom-specific approach to
treatment, targeting fatigue and other residual symptoms
that may affect long-term outcomes (Fava, 2003, 2006;
Kennedy, 2008; Kurian et al., 2009; Menza et al., 2003; Pae
et al., 2007; Papakostas et al., 2006). Clinical trials of
symptom-specific treatments will require measures that can
capture these individual symptoms and their impact. Thus,
the FAsD may be helpful in the development and evaluation
of treatment interventions focusing specifically on fatigue
associated with depression.

Role of funding source
Funding for this study was provided by Eli Lilly and Company,
Indianapolis, IN, USA. The study sponsor played no role in the study design,
data collection, data analysis, interpretation of data, writing of the report, or
the decision to submit the paper. One of the authors (Glenn Phillips) was an
employee of Lilly, but his input into the conceptualization of this study
represented his own opinions, rather than those of the company.

Appendix A. Fatigue Associated with Depression Questionnaire

Some people experience fatigue when they are depressed. The following items ask you to rate fatigue you have experienced
that you think may be related to depression.

Please mark an “X” in the box that best describes your experience during the past week.

<table>
<thead>
<tr>
<th>In the PAST WEEK, how often have you felt...</th>
<th>Never (1)</th>
<th>Rarely (2)</th>
<th>Sometimes (3)</th>
<th>Often (4)</th>
<th>Always (5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Fatigued</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Tired</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Exhausted</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Like you had no energy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Physically weak</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Like everything requires too much effort</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Now think about the impact of this fatigue that is related to depression. The following items ask about the impact of this fatigue
on various aspects of your life.

<table>
<thead>
<tr>
<th>In the PAST WEEK, how much has your fatigue...</th>
<th>Not at all (1)</th>
<th>A little (2)</th>
<th>Somewhat (3)</th>
<th>Quite a bit (4)</th>
<th>Very much (5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>7. Limited your ability to complete daily household chores</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Interfered with family activities or relationships</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Interfered with doing things you enjoy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Interfered with social activities, like spending time with friends</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. Interfered with taking care of yourself (e.g., bathing, dressing, brushing your teeth)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Do you have a spouse or significant other?
☐ Yes (Please answer item 12.)
☐ No (Leave item 12 blank.)

<table>
<thead>
<tr>
<th>In the PAST WEEK, how much has your fatigue...</th>
<th>Not at all (1)</th>
<th>A little (2)</th>
<th>Somewhat (3)</th>
<th>Quite a bit (4)</th>
<th>Very much (5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>12. Interfered with your intimate relationship (i.e., with a spouse or significant other)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Do you have a job or go to school?
☐ Yes (Please answer item 13.)
☐ No (Leave item 13 blank.)

<table>
<thead>
<tr>
<th>In the PAST WEEK, how much has your fatigue...</th>
<th>Not at all (1)</th>
<th>A little (2)</th>
<th>Somewhat (3)</th>
<th>Quite a bit (4)</th>
<th>Very much (5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>13. Limited your productivity at work or school</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
References


